



## REFERENCE ONLY

### UNIVERSITY OF LONDON THESIS

Degree **PhD**

Year **2006**

Name of Author **TOPINTZ, E**

#### COPYRIGHT

This is a thesis accepted for a Higher Degree of the University of London. It is an unpublished typescript and the copyright is held by the author. All persons consulting the thesis must read and abide by the Copyright Declaration below.

#### COPYRIGHT DECLARATION

I recognise that the copyright of the above-described thesis rests with the author and that no quotation from it or information derived from it may be published without the prior written consent of the author.

#### LOANS

Theses may not be lent to individuals, but the Senate House Library may lend a copy to approved libraries within the United Kingdom, for consultation solely on the premises of those libraries. Application should be made to: Inter-Library Loans, Senate House Library, Senate House, Malet Street, London WC1E 7HU.

#### REPRODUCTION

University of London theses may not be reproduced without explicit written permission from the Senate House Library. Enquiries should be addressed to the Theses Section of the Library. Regulations concerning reproduction vary according to the date of acceptance of the thesis and are listed below as guidelines.

- A. Before 1962. Permission granted only upon the prior written consent of the author. (The Senate House Library will provide addresses where possible).
- B. 1962 - 1974. In many cases the author has agreed to permit copying upon completion of a Copyright Declaration.
- C. 1975 - 1988. Most theses may be copied upon completion of a Copyright Declaration.
- D. 1989 onwards. Most theses may be copied.

***This thesis comes within category D.***

☒

This copy has been deposited in the Library of UCL

☐

This copy has been deposited in the Senate House Library, Senate House, Malet Street, London WC1E 7HU.



# **Moraic Onsets**

**Ioanna Topintzi**

Department of Phonetics and Linguistics  
University College London

Thesis submitted to the University of London in partial fulfilment of  
the requirements for the degree of Doctor of Philosophy

**June 2006**

UMI Number: U592441

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U592441

Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.  
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against  
unauthorized copying under Title 17, United States Code.



ProQuest LLC  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106-1346



## **Signed declaration**

This dissertation is the result of my own work, unless stated otherwise in the text or footnotes.

Ioanna Topintzi

## Abstract

This thesis examines the status of onsets and their effects on stress and prosody. I argue that moraic onsets exist, a claim that contradicts standard phonological models (Hyman 1985, Hayes 1989, Gordon 1999, Morén 1999) which assume that onsets are not moraic, given that in the overwhelming majority of languages onsets are inert for prosodic processes.

Using data from Pirahã, Karo and Arabela stress, I show that weightful onsets actively participate in weight-sensitive stress assignment. Moreover, I point out that if onsets can be moraic, a host of other weight-based phenomena, should also be able to utilize them. This is exactly right, as verified by word minimality in Bella Coola, Samothraki Greek compensatory lengthening (CL), onset geminates in Pattani Malay, Trukese and Marshallese and a variety of other data, e.g. Trique CL, Bellonese reduplication. Crucially, this is not a prediction shared by previous prominence-based analyses of similar facts (Hayes 1995, Gordon 2005, Smith 2005). Prominence is inherently designed to account only for stress, not for other weight-based phenomena. If one were to entertain a prominence account, then most of the data above would remain unexplained.

However, not all onsets can be moraic. The proposed model is restrictive in admitting only two kinds of moraic onsets: those which are underlying, i.e. emerging as geminates, and those which are derived in the output and serve for stress purposes. While the former can be of any featural content (since they are lexically specified and thus unpredictable), for the latter ones, I claim that only voiceless onsets can be moraic, whereas voiced ones are never moraic. This relates to a well-known generalisation affecting a different prosodic phenomenon, namely tone. Voiceless onsets raise the pitch of the following vowel, voiced ones lower it. In many languages such pitch perturbation is interpreted as tone. My proposal is that in some other languages, this pitch perturbation is instead interpreted as stress and is formally represented by means of moras, which are only assigned to the stress-attracting voiceless onsets. Pirahã, Karo and Arabela data empirically confirm this finding.

## Acknowledgements

Rarely things work ideally in life, but having Moira Yip as a supervisor certainly has. Moira's fascinating knowledge of just about everything phonological, her expertise and her incredibly quick mind are truly remarkable. I have benefited immeasurably from her sharp comments and from challenging my ideas and thoughts from day one. Thanks for devoting so much time and energy to me. I hope this thesis can somehow compensate for both. It has been a privilege to work with you and for being allowed to see, as time progressed, what a truly alluring person you are outside of linguistics too.

Many people have also helped me by providing suggestions, comments or making papers and material available to me. I gratefully acknowledge: Juliette Blevins, Lev Blumenfeld, Emily Curtis, Mark Donohue, Paloma Garcia-Bellido, Spike Gildea, Rob Goedemans, Carolina González, Beverley Goodman, Matt Gordon, Carlos Gussenhoven, John Hajek, Bruce Hayes, Ben Hermans, Wouter Jansen, Brian Jose, Dasha Kavitskaya, Yen-Hwei Lin, Bruce Morén, Elliott Moreton, David Nash, Roland Rich, Curt Rice, Keren Rice, Jen Smith, Paul Smolensky, Donca Steriade, Marc van Oostendorp, Yi Xu and Hideki Zamma. Special thanks to Mary Ruth Wise for bringing Arabela to my attention and Debbie Koop at SIL Peru for making relevant material accessible to me. Thanks to Eleftheria Giakoumaki and Angelos Afroudakis at the Κέντρον Ερεΐνης των Νεοελληνικών Διαλέκτων και Ιδιωμάτων - I.A.N.E for granting me access to older Samothraki Greek sources. Dan Everett's work on Pirahã has inspired me to write a thesis on this topic. Thanks for answering all my - occasionally silly - questions about the language in lightning speed.

At UCL, thanks to my fellow-students for making life at the PhD room, and sporadically outside of it, enjoyable: Nick Allott, Eric Carlson, Ana Carrera-Hernandez, Alison Hall, Vikki Janke, Eirini Sanoudaki, Hitoshi Shiraki, Marco Tamburelli, Rob Truswell, Hiroyuki Uchida, Rosa Vega-Moreno and Reiko Vermeulen. Special thanks to Steve Nevard and those of you who happily offered to participate as subjects in my experiment, to Molly Bennett for being so helpful and efficient in all things practical, to Stavroula Kousta for invaluable help regarding statistics, to Gloria Malambe and Mary Pearce for their friendship, but also to Mary for collaboration in phonology and editing. I am grateful to Neil Smith and Valerie Hazan for being there to assist with any practical or funding problems, and to John Harris for being keen to discuss my work.

My stay in England and the completion of my thesis have been greatly facilitated by grants from the AHRC, the A.G. Leventis Foundation, the British Federation of Women Graduates and the UCL Friends. Thanks to the UCL Graduate School and the Department of Phonetics and Linguistics for partially funding my conference trips. I gratefully acknowledge their financial support.

To the following fellow Greek linguists: Elena Anagnostopoulou, Mary Baltazani, Ioanna Kappa, Marika Lekakou, Angeliki Malikouti-Drachman, Marianna Margariti-Ronga, Dimitris Papazachariou, Angeliki Ralli, Anthi Revithiadou and Marina Tzakosta, σας ευχαριστώ πολύ όλους for your help, comments and friendship.

On a more personal level, I wish to thank Marika Lekakou for being a close companion over the past two years. I am grateful to you for sharing thoughts, for liking the fact that I can be 'difficult', for discussions about life and linguistics, but above all for a cucumber eye-treatment I will never forget.

Choosing to do linguistics in the first place was Elena Anagnostopoulou's 'fault'. Thank you for that. And thanks for teaching me syntax and for forgiving me for not doing syntax. Your trust in me from the very beginning as well as your friendship, have always been precious.

Στον Μαρκ από το ανατολικό χωριό, thanks for your drawings, your baby Greek and for all the things you know and more. We are still on the path, and you know it too.

To my Greek friends in London: Stavros Fatouros, Elena Giouroukou, Georgia Iliopoulou, Eftihia Karadimitri, Konstantinos Lasithiotakis, Dimitris Makris (aka Jim Long), Tania Vasileiadou, thank you all and each and every one for enriching my life in so many ways, for challenging me to try new things, for the best Easter ever, for overeating evenings in Enfield pubs and Diyarbakir and for singing τα τραγούδια της παρέας every now and then. Thanks to all my friends in Greece, and particularly to my lifelong friends: Elisavet Kokkoni, Giouli Kommata, and 'my brother' Kostas Simatos, for only needing a split second to 'warm up' every time we meet and for meeting up every time I'm there. To Konstantina Hourdaki, thanks for being my biennial 'therapist' and for listening. Hope I can pay you back sometime.

There is so much I owe to my mum Betty and dad Takis and so little space to list them all. I can never thank you enough for your unfailing and unconditional love, emotional and financial support and for occasionally making the effort to understand what it is I'm doing (let alone try to describe it to others...). Many thanks to my aunt Katerina, uncle Pavlos and cousin Eleni for making family gatherings an event I never want to miss.

A few years ago I would laugh at the idea that my sister Ermina would become my best friend, most ardent supporter and funniest housemate. But I am so glad that this became a reality. I owe you eternal thanks for everything. And as this will take forever, I'll do something more realistic and dedicate this thesis to you.

# Contents

<b>Abstract</b>	<b>3</b>
<b>Acknowledgements</b>	<b>4</b>
<b>Chapter 1 Theoretical Issues</b>	<b>11</b>
1.1 Syllable weight	11
1.2 Types of onset sensitive stress	16
1.3 Previous analyses	19
1.3.1 Prominence-based accounts	20
1.3.2 Smith (2005)	24
1.3.3 Gordon (2005)	28
1.4 Onset moraicity	32
1.4.1 Types of moraic onsets: distinctive vs. coerced weight	32
1.4.1.1 The effect of voicing on tone and stress and its consequences for onset moraicity	34
1.4.1.2 Generating moraic onsets in coerced weight	44
1.4.1.3 Moraic onset geminates	46
1.5 Organisation of the thesis	51
<b>Chapter 2 Onset Moraicity and Stress</b>	<b>53</b>
2.1 Introduction	53
2.2 Pirahã	56
2.2.1 Data and generalisations	56
2.2.2 Analysis	58
2.2.3 An alternative: Gordon (2005)	67
2.3 Arabela	73
2.3.1 Data and generalisations	73
2.3.2 Analysis of the stress system	77
2.3.2.1 The basic analysis	77
2.3.2.2 The importance of equal ranking	81
2.3.2.3 Testing the system: voiceless-sonorant combinations and non- shift of stress in other positions	85



4.2.2.1 Root Maximality	148
4.2.2.2 Word Minimality	153
4.2.2.3 Problems for Word Minimality in Bagemihl (1998) and a solution	155
4.3 The core analysis	157
4.4 Detailed analysis	163
4.4.1 Word Minimality	163
4.4.2 Against moraic onsets	170
4.4.3 The remaining roots	171
4.4.3.1 Well-formed roots	172
4.4.3.2 Larger roots and root maximality	173
4.4.4 Summary and Final Ranking	176
4.5 Residual issues	179
4.5.1 Morphological Root vs. Morphological Word	179
4.5.2 MPARSE	181
4.6 Conclusion	184
<b>Chapter 5 CL from onset loss: the case of Samothraki Greek</b>	<b>185</b>
5.1 Introduction	185
5.2 Hayes' (1989) model and problems for OT	186
5.3 CL as position preservation	189
5.4 Samothraki Greek	196
5.4.1 The data	196
5.4.1.1 Summary of the patterns	200
5.4.2 The analysis	200
5.4.2.1 On the placelessness of SamG /r/ and constraints required	200
5.4.2.2 Coda /r/	202
5.4.2.3 Singleton /r/ word-initially	202
5.4.2.4 Singleton /r/ word-medially	203
5.4.2.5 Complex clusters of the type Cr+V+C - the simple(r) case	204
5.4.2.6 Complex clusters of the type Cr+i/e+V - the difficult case	204
5.5 Alternatives to CL as position preservation	208
5.5.1 Hayes (1989)	209
5.5.2 OT accounts	210
5.5.2.1 Intermediate levels	210
5.5.2.2 Segmental faithfulness	212
5.5.2.3 CL as coalescence (Sumner 1999)	214
5.5.3 A phonetically-based approach - Kavitskaya (2002)	217

5.6 An extension: CL in Piro	219
5.7 On the rarity of CL after onset loss	224
5.8 Conclusion	227
<b>Chapter 6 Moraic onsets and geminates</b>	<b>230</b>
6.1 Introduction	230
6.2 Theoretical background	231
6.3 Initial moraic geminates as moraic onsets ('Initial' geminates word-initially)	233
6.3.1 Pattani Malay and stress	233
6.3.1.1 The data	233
6.3.1.2 Hajek and Goedemans (2003)	237
6.3.1.3 Current analysis and the role of vowels	241
6.3.1.4 Pattani Malay summary	249
6.3.2 Trukese Word Minimality	249
6.4 Moraic onsets in medial geminates ('Initial' geminates word-medially)	258
6.4.1 Theoretical background	258
6.4.2 Marshallese	260
6.5 Remaining issues	265
6.5.1 Geminate weight and length	265
6.5.2 'Weightless geminates'	271
6.6 Conclusion	273
<b>Chapter 7 Ambiguous and misanalysed onset-sensitive languages</b>	<b>274</b>
7.1 Introduction	274
7.2 Languages suggestive of onset moraicity subject to satisfaction of certain conditions	275
7.2.1 Geminates and reduplication	275
7.2.1.1 Bellonese	275
7.2.2 CL	277
7.2.2.1 Onondaga	277
7.2.2.2 Proto Austronesian	278
7.2.2.3 Trique	279
7.2.3 Word Minimality	281
7.2.3.1 Damin	281
7.2.4 Metrics	283
7.2.4.1 Luganda court songs	283
7.2.5 Remaining cases	284
7.2.5.1 Supyire	284



7.2.5.2 Yoruba	287
7.2.5.3 Tukang Besi	289
7.3 Misanalysed onset sensitive languages	291
7.3.1 Berawan	291
7.3.2 Bislama	297
7.3.3 Nankina	301
<b>Chapter 8 Concluding remarks and directions for future research</b>	<b>305</b>
8.1 Concluding remarks	305
8.2 Experiment	309
8.3 Directions for future research	320
8.3.1 Moraic contrast between CV and CCV?	321
8.3.2 Onsets as TBUs?	322
8.3.3 Language documentation and attention to the role of onsets	329
<b>Appendix A: Full tableaux for Bella Coola [Ch. 4]</b>	<b>332</b>
<b>Appendix B: Bislama patterns according to Camden (1977) [Ch. 7]</b>	<b>335</b>
<b>References</b>	<b>336</b>

## Chapter 1

### Theoretical issues

*“... no language has a rule stressing the penultimate syllable unless it begins with a voiced consonant, in which case one stresses the antepenultimate syllable”*  
(Hyman 1985: 96)

*“(Karo) stress can be predicted by the onset of the last syllable: if it is a voiced stop consonant, then the stress shifts one syllable to the left”*  
(Gabas 1999: 39)

#### 1.1 Syllable weight

The notion of syllable weight is pervasive in phonology. Everywhere we look there are data that make reference to it. The idea behind syllable weight is that depending on their structure, syllables are treated in a different way; thus if a syllable contains a long vowel or - in some languages - a short vowel followed by a coda consonant, then it is heavier than one including a short vowel, i.e. VV, VC > V. The effects of this distinction are most prominently seen in stress, where in many languages heavy syllables attract stress more than light ones. Importantly, VC is heavy in some languages, e.g. Hopi, but light in others, e.g. Lenakel (see below).

In Hopi (Gussenhoven and Jacobs 2005: 145), the first syllable is stressed if it is heavy (C)VV or (C)VC [(1a)], but if it is (C)V light, then the second gets stress [(1b)]<sup>1</sup>.

(1) *Hopi: VV/VC=heavy; V=light*

- |    |               |                  |
|----|---------------|------------------|
| a. | qóq.tø.som.pi | ‘headbands’      |
|    | só:.ja        | ‘planting stick’ |
| b. | qø.tø.som.pi  | ‘headband’       |
|    | ko.jó.ŋo      | ‘turkey’         |

---

<sup>1</sup> In this chapter, *unless stated otherwise*, the acute accent marks primary stress, the grave accent means secondary stress, and underlining denotes the reduplicated portion. I will interchangeably use VV or V: to refer to long bimoraic vowels.

In Lenakel (Hayes 1995) on the other hand, (C)VCs are considered light and primary stress appears on the penult (2a). Simplifying a bit, secondary stress is (usually) assigned on the first syllable and on every other syllable after that (2b). This pattern however can be disrupted; while a (C)VC will not get secondary stress unless it happens to be located in a position that would receive rhythmic stress anyway (cf. unstressed *mol* in (2d)), heavy (C)VV gets stress no matter what its position (cf. *kì:* in 2c).

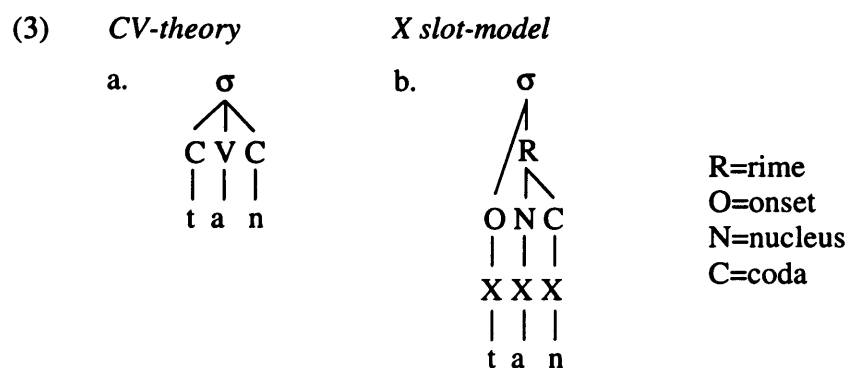
(2) *Lenakel: VV=heavy; VC/V=light*

- |    |                             |                      |
|----|-----------------------------|----------------------|
| a. | éheŋ                        | ‘to blow the nose’   |
|    | rimáwŋin                    | ‘he ate’             |
| b. | lètup <sup>w</sup> álukáluk | ‘in the lungs’       |
| c. | níkì:nílar                  | ‘their (pl.) hearts’ |
| d. | řimolkéykey                 | ‘he liked it’        |

Several other phenomena are sensitive to syllable weight. A by-no-means-exhaustive list includes: i) compensatory lengthening, i.e. the lengthening that occurs after a segment’s deletion, e.g. Turkish (Roca and Johnson 1999) *tahsil* → *ta:sil* ‘education’; ii) word minimality, that is, the minimum word size some languages impose [commonly (C)VC or (C)VV], e.g. Dalabon (Capell 1962, Garrett 1999) allows words that are CVC *bad* ‘stone’ or CVV *bi:* ‘man’, but no CVs; iii) metrics, that is the organization of syllables into feet in songs or poetry. For instance, the most prominent meter in Greek and Latin epic poetry is the dactylic hexameter, where the verse consists of six metra, each of which is made of one heavy and two short syllables (– ∪ ∪); however two short syllables can be replaced by one heavy in which case we have a spondee (– –). In the example below boundaries of metra are marked by parentheses: (cārmīnǎ) (quāe vūl)(tīs cōg)(nōscītē); (cārmīnǎ) (vōbīs) [Vergil, *Eclogues* VI. 25]; iv) reduplication, i.e. the repetition of part of a word that commonly needs to have the size of a heavy syllable, e.g. Mokilese progressive (Harrison 1976, McCarthy and Prince 1986) *poki~pokpoki* ‘beat’, *kookɔ~kookkookɔ* ‘grind coconut’, but *pa~paapa* ‘weave’; v) (prosodic) truncation as in nickname formation, e.g. among other patterns, acceptable Japanese nicknames are a single heavy syllable (Mester 1990, Benua 1995) *Midori~Mii-čan* or *JuNko~JuN-čan* (–čan is the diminutive suffix); vi) gemination, e.g. the consonant doubling that often occurs after short stressed vowels. Creation of a heavy syllable in this manner can serve stress purposes, since some languages require that all their stressed syllables are heavy (Italian). Gemination of this type occurs in Kukatj (Breen 1992) or in Swedish dialects (Kiparsky to appear) such as *viss.na* ‘to wilt’, *takk.sa*

‘rate’, *hall.va* ‘half’; vii) tone. In particular, contour tones may frequently only be tolerated in heavy, but not in light CV syllables. Hausa (Gordon 1999) is a representative example of this sort, e.g. *lâ:lâ:* ‘indolence’, *mântá:* ‘forget’, *râs:â:* ‘branches’.

The distinction between heavy and light syllables is recognised as early as Jakobson (1931) and Trubetzkoy (1939) and has since been formalised in three major ways: a) CV theory (McCarthy 1979, Clements and Keyser 1983), b) X slot model (Levin 1985), c) moraic model (Hyman 1985, Hayes 1989). All three theories assign abstract weight units to segments in the syllable. The difference lies in what kinds of units these are and what exact syllable constituents are recognised, which of course has repercussions on the predictions made. For example, the syllable *tan* would be represented in the first two models in the following way.



In CV theory (3a), segments are distinguished as C-ones and V-ones. A welcome result is that by doing so, syllabic segments (e.g. syllabic sonorants, vowels) are separated from non-syllabic ones. For instance, a sonorant that is linked to a C is non-syllabic, but when linked to a V, it is syllabic. Sometimes however, the C and V label is far too specific. For instance, in Ancient Greek, the form *esmi* ‘I am’ underwent *s*-deletion and subsequent compensatory lengthening. In some dialects, the resulting form was *emmi* with C-lengthening, while in others it was *eemi* with V-lengthening. CV-theory can account for *emmi*, because the vacated C-position of *s* is filled by a consonant too, but it fails to do so in *eemi* where the position is held by a V.

This is not a problem that X-slot theories face (3b), since X slots, by being general enough, circumvent this problem of labelling. An issue that arises though is that X-slot models suggest that counting of X units is a sufficient mechanism to account for e.g. reduplication templates. But this commonly proves wrong. For instance, in Mokilese the reduplicated progressive form for /diar/ ‘find’ is [dii-diar] and not

[dia-diar], as an X-slot theory would predict, if filling three timing slots was all that was required.

The advent of moraic theory (Hyman 1985 and particularly Hayes 1989) aims at solving these problems as well as addressing an issue which arises in both previous theories and which is central to this thesis. Notice that both of these assign a timing slot to the onset and thus seem to allow for the possibility that this constituent may be active in prosodic processes too. However, theorists of the moraic framework claim that this is never the case, consequently, the model must be modified to directly mirror this fact.

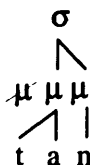
Hyman (1985) proposes a model which consists of weight units (WU's) whose function is virtually identical to moras, which is why I will simplify and use moras for Hyman's representations too (4a). For our purposes, the most important property of this model is that underlyingly all segments start off with at least one WU (4a.i). Crucially, on the surface onsets lose their WU (indicated by the crossed-out mora in (4a.ii)) due to the universal application of the Onset-Creation Rule (OCR). This rule applies whenever a [+cons] segment is followed by a [-cons] segment and its effect is to delete the WU of the [+cons] segment. Subsequently, the [+cons] feature matrix associates to the WU of the [-cons] segment on its right. In other words, the nucleic WU/mora dominates both the onset and the core of the syllable (4a.ii).

(4) a. *Hyman (1985)*

i. underlying form



ii. surface form



b. *Hayes (1989)*

surface form



This differs from Hayes (1989) who assumes that the nucleic mora is not shared between the onset and the nucleus, but only associates to the nucleus. The onset instead links directly to the syllable node as depicted in (4b). Note that although I have represented the coda consonant in (4b) as moraic, a singleton coda consonant may be non-moraic on the surface (compare Hopi (C)V<sup>μ</sup>C<sup>μ</sup> with Lenakel (C)V<sup>μ</sup>C above). If it is moraic, this is the result of the application of the Weight-by-Position rule which assigns moras on codas.

Moraic theory has several advantages over the previous models. First, it captures the syllabic status simply through the presence of a mora; second, it allows processes such as reduplication to merely count moras and let other syllable requirements regulate

the exact shape of the reduplicant. Finally, problems of the *esmi* type above no longer arise. Compensatory lengthening (CL) is just about mora preservation (but also see Chapter 5). Either V- or C-lengthening can fulfil this goal and both are in principle available. Thus, in our *esmi* example from Ancient Greek, dialectal variation between *eemi* and *emmi* as a product of CL, is natural.

With moraic theory, it has therefore been possible to account for a broad set of facts in a uniform manner. Heavy syllables are merely the ones that consist of two moras, whereas light syllables comprise only one. Moras are grouped into feet (McCarthy and Prince 1986, Hayes 1995), which are part of higher prosodic structure that includes prosodic words (Selkirk 1980, 1984; Nespor and Vogel 1986; Itô and Mester 1992). Reference to feet and moras allows us to account for numerous data, many of which cannot be adequately accounted for in other timing models.

For instance, in Lardil (McCarthy and Prince 1986), words must minimally consist of two moras, i.e. a foot. Only vowels count as moras. Words that contain a single mora augment by suffixation of a morphologically empty *a* (5b). Words that already satisfy minimality undergo no change. These can either include a long vowel or consist of two light syllables (5a), a pattern captured in terms of moras, but missed when counting X-slots (or C/V slots).

(5) *Lardil augmentation achieves bimoraicity*

	Underlying	Uninflected	
a.	/peer/	peer	'ti-tree sp.'
	/kela/	kela	'beach'
	/parŋa/	parŋa	'stone'
b.	/wik/	wika	'shade'

Similarly, recall that Japanese nickname formation creates hypocoristics such as *Midori~Mii-čan* or *JuNko~JuN-čan*. These can be simply viewed as heavy CVV or CVC syllables, a fact that can be expressed in all the theories above. However, for other names, various possible nicknames are available.

(6) *Japanese Hypocoristics* (Benua 1995)

Midori	Mido-čan, Mii-čan
Hanako	Hana-čan, Haa-čan, Hač-čan

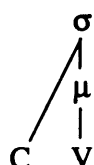
While these hypocoristics can receive a uniform and straightforward account in a moraic framework, i.e. they are a bimoraic foot whether this implies monosyllabic

heavy CVV/CVC or light CV disyllables, no similar generalisation obtains in the other models. C/V- or X-slots would need to impose three or four slots depending on the nickname considered each time, but of course the real insight behind this process is lost.

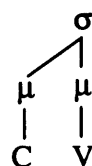
For all these reasons, moraic theory has proven more successful compared to its predecessors, and because of this, it will be the timing model assumed in this work. Importantly however, moraic theory à la Hayes (1989) claims that an additional strong point of the model is that it does not allow moraic onsets. Hayes argues that onset consonant deletion never causes CL, which is only natural if onsets never carry weight. Notably, this effect cannot be derived in the previous frameworks, at least not without the introduction of the rimal node which is the only one that can bear weight.

As we will see in what follows, this is exactly the point I question in this thesis. I will claim that moraic onsets *do* exist in some languages and are represented in the way shown in (7b). Their introduction does not undermine moraic theory, but aspires to improve the range of facts that moraic theory can account for; in fact, any ban on moraic onsets is purely stipulatory, as I will argue below. I am not suggesting that the presence of moraic onsets is unrestricted, but that it is regulated by certain patterns pertaining to voicing (§1.4.1.2) or underlying moraic specification. Under this view, moraic theory remains in an advantageous position, because even after the introduction of moraic onsets, it can still distinguish between languages that have them (7b) versus the ones that do not (7a) by simply assigning a mora on the onset of the first, but not of the latter.

(7) a. *Non-moraic onsets*



b. *Moraic onsets*



Other timing models do not have this option; onsets should either be consistently moraic or non-moraic across languages. Neither situation though reflects reality.

## 1.2 Types of onset sensitive stress

As we have seen, syllable weight plays a pivotal role in phonology. However, the term syllable weight itself is misleading, since it is standardly assumed that syllable weight is really defined as the rimal weight only and excludes the onset. “Onset segments are prosodically inert... While this claim is not fully valid at the observational level, it is so

well supported across languages that it serves as the central observation for formal theories of syllable weight” (Hayes 1995: 51). As Hayes admits, empirical data from a few languages suggest that at least at the *observational level* onsets can in fact be prosodically active. It is of course true that there is a very strong tendency that onsets do not matter for weight purposes, but this can by no means be a universal.

Similar statements are made by other researchers too, who include this small caveat, namely that it is not absolute that onsets never ever play any role to weight. For example, Gordon (1999: 3) observes that “in Latin, as in *virtually* all languages, the onset is ignored for purposes of calculating weight” (emphasis added mine). In a similar vein, Morén (1999/2001: 7-8) states that “...onsets are typically non-moraic. Although this is not the only logical possibility, it is *convenient* and I assume it here” (emphasis added mine).

The present thesis will in fact deal with this “inconvenient” issue and will attempt to show that certain stress and syllable weight facts cannot be re-analysed in any way other than by admitting weightful onsets. This suggests a more literal interpretation of the term ‘syllable weight’. Not only does syllable weight refer to rimal weight, but also to onset weight. I will thus argue for the existence of moraic onsets, a rather controversial claim in the light of so much work arguing against it, including most notably Hyman (1985), Hayes (1989), Morén (1999/2001), Gordon (1999) and many others.

These studies are based on the preponderance of phonological phenomena and processes that ignore the onset for weight purposes. In fact, it has been proposed that one of the significant advantages of moraic theory – at least as advanced in Hayes (1989) – over CV- or X-theories, is that it can capture the onset-rime distinction without any stipulation by simply assigning moras to the rime portion and leaving the onset moraleless. Thus, in standard moraic theory, the onset is defined as the prenucleic non-moraic consonant. However, this is really a stipulation too, a ‘convenient assumption’ as Morén (1999/2001) puts it. In principle, there is nothing wrong with having a moraic onset. Consequently, within the current proposal, onsets come into two flavours: non-moraic (7a) - as in most languages - and moraic (7b). An onset can still be seen though as the tautosyllabic prenucleic consonant.

Focusing on stress in particular, there is a small number of languages whose stress algorithm needs to take onsets into account. In fact, I argue that there are three major types of languages where onset effects are relevant for stress: those that are sensitive to the presence of the onset; those that are sensitive to the type of the onset,



and finally those that are sensitive to both. It is my claim that other researchers (e.g. Buller, Buller and Everett 1993; Davis 1985, 1988; Everett 1988; Goedemans 1996, 1997, 1998; Gordon 2005) who have worked on this issue have implicitly or explicitly conflated these three types of languages without any further distinctions.

I instead attempt to show that the effect the presence or absence of an onset may have on stress is a different phenomenon from that relating to the type of the onsets involved. This makes an extra prediction. If we are talking about two distinct phenomena, then it is likely that they may interact with one another. This expectation is borne out, as shown in (8). Some languages show effects attributed to the presence of an onset (❷), some to its quality (❸), some are sensitive to both (❶), while others to none (❹). This yields a four-way typology:

(8) *Presence and quality of onset interaction in stress*

		Quality of onset	
		YES	NO
Presence of an onset	YES	Pirahã (Arabela) <sup>2</sup> (❶)	Aranda, Banawá, Dutch (❷)
	NO	Karo (Arabela) (❸)	Standard Greek, Russian, etc. (❹)

In Pirahã, the rightmost heaviest syllable of the final three in a word receives stress according to the following hierarchy: PVV > BVV > VV > PV > BV (P = voiceless onset, B = voiced onset, > = is heavier than). Thus, stress is final if all syllables are of the same type (9a), but can appear elsewhere when syllables are different. In particular, onsetful syllables attract stress more than onsetless ones (9b; gáí > aí), but also onsets of a certain type, i.e. voiceless obstruents, attract stress more than the rest (9c; ʔí > bo, gi).

(9) *Pirahã* (stressed syllables in bold; acute accent = H tone, no accent = L tone)

- a. ko.ʔo.**pa** “stomach”
- b. poo.**gáí**.hi.aí “banana”
- c. ʔí.bo.gi “milk”

In Karo, default stress is word-final, unless some requirement, i.e. tone, nasalization or onset voicing, causes shift from that position. In particular, final voiced obstruent onsets repel stress (10a), whereas voiceless obstruent (10b) and sonorant (10c) onsets do not.

<sup>2</sup> Arabela most likely lacks onsetless syllables, therefore it cannot serve as a testing ground with respect to the presence of the onset issue. This is why I position it in both cells.

Onsetless syllables are allowed (Gabas 1999: 24), and they can carry stress too if they make the best available stress bearers (10d). This suggests that onsetless syllables in Karo are not treated in any special way.

(10) *Karo onsets and stress* (Gabas 1999: 39-41; stress in bold)

- |    |                 |              |                  |
|----|-----------------|--------------|------------------|
| a. | <b>cigi</b>     | ‘spot’       |                  |
|    | <b>pibɛʔ</b>    | ‘foot’       |                  |
| b. | <b>pakɪɔ</b>    | ‘fish (sp.)’ |                  |
|    | <b>nahekʔ</b>   | ‘fontanel’   |                  |
| c. | <b>maŋʔɪtʔ</b>  | ‘again’      |                  |
|    | <b>kiriwɛpʔ</b> | ‘butterfly’  |                  |
| d. | <b>pɛ.ɔʔn</b>   | ‘skin’       |                  |
|    | <b>e.i</b>      | ‘irara’      | (Gabas 1998: 22) |

In Aranda on the other hand, stress on onsetless syllables is avoided, so that actually the first onsetful syllable receives stress irrespective of its type (compare onsetful (11a) with onsetless (11b)). Finally, in languages such as Greek, the presence or quality of onsets plays absolutely no role in the stress algorithm. Syllables with onsets of any type may receive stress (12a) and onsetless syllables may carry stress too (12b).

(11) *Aranda: onsets and stress* (Strehlow 1944; diacritics ignored)

- |    |        |              |             |                        |
|----|--------|--------------|-------------|------------------------|
| a. | tárama | ‘to laugh’   | kútuŋùla    | ‘ceremonial assistant’ |
| b. | ankáta | ‘Jew lizard’ | ulámbulàmba | ‘water-fowl’           |

(12) *Greek: no effect from onsets on stress*

- |    |         |                 |        |        |
|----|---------|-----------------|--------|--------|
| a. | pérazma | ‘way-through’   | perúka | ‘wig’  |
| b. | étimos  | ‘ready-MASC-SG’ | eláfi  | ‘deer’ |

In the next section, I will consider the most interesting previous accounts of these facts and discuss the drawbacks they present. Subsequently (§1.4), I will offer the alternative offered in the current work.

### 1.3 Previous analyses

There are three lines of reasoning to account for the facts above, particularly those pertaining to pattern ❷ and the ‘quality of onset’ part of pattern ❶<sup>3</sup>. The most popular involves some notion of prominence (e.g. Everett 1988, Hayes 1995, Smith 2005).

<sup>3</sup> Pattern ❷ that refers to the presence of an onset is treated as a different phenomenon, i.e. as an alignment effect where the stressed syllable needs to be aligned with an onset. More on this pattern can be found in Chapter 3. This is why at present, my discussion and criticism of earlier accounts for the most part centres around the ‘quality of onset’ issue.

Another recent proposal (Gordon 2005) makes use of both prominence and weight, whereas the current account makes use of an exclusively weight-based account. In what follows, I will first review some prominence-based accounts and point out certain serious shortcomings (§1.3.1). Next, I will discuss the most successful of these accounts (Smith 2005) in more detail (§1.3.2). I will then move on to consider Gordon's (2005) proposal that integrates both prominence and weight and show that this is not flawless either (§1.3.3). At that point, I will only stick to some general comments, and return to some further criticism and technical problems for Gordon in §2.2.3.

### 1.3.1 Prominence-based accounts

Like syllable weight, prominence is another notion widely used in phonology. Unlike syllable weight though, whose manifestation is more clearly identified, the definition of prominence is elusive. Jensen (2004) correctly observes that depending on the analyst, prominence has different meanings and is sometimes related to - although is considered distinct from - akin notions like *stress* and *accent*. It is therefore generally agreed, citing Jones (1909: 141), that "Stress is not the same as 'prominence' [...]; stress is one of the factors that may cause or help to cause a sound or syllable to be 'prominent'". According to Jones (1909: 142), the other factors which can make a sound more prominent are "inherent sonority, length and intonation". Despite the lack of consensus on what prominence entails exactly, a shared assumption in more recent work seems to be that prominence is equated with "perceptual salience" (Hayes 1995, Jensen 2004, Smith 2005), which is again quite vague.

In spite of this weakness, prominence has been used widely to account for several data. More specifically for Pirahã (9), which is of interest to us, analyses such as Everett (1988) and Hayes (1995) have been proposed that use an arbitrary prominence grid that renders voiceless obstruents more prominent and thus more stress attracting than their voiced counterparts. Hayes' account of Pirahã follows:

(13) *Hayes' (1995: 286) Pirahã prominence scale*

[P = voiceless stop, B = voiced stop, > = is heavier than]

A. Prominence Projection: Project prominence grids as follows:

\*\*\*\*\* : PVV  
 \*\*\*\* : BVV  
 \*\*\* : VV  
 \*\* : PV  
 \* : BV

B. Apply End Rule Right within the final trisyllabic domain.

While the Pirahã facts are accounted for by the scale in (13), there is no explanation why such a scale should hold. As a matter of fact, in principle, nothing precludes a scale where for instance BVV has more prominence grids than PVV or BV more than PV. This is entirely unaccounted for. Goedemans (1998) attempts to explain this direction of the facts in a more principled manner phonetically, but eventually does not go on to propose a more fully-fledged analysis. This task is taken up by Gordon (2005) whose approach also hinges on phonetic considerations and the notion of prominence, but has specific phonological proposals too. On the contrary, as we will see next in more detail, the current approach attributes this distribution to the effect of voicing or lack thereof on pitch perturbation and suggests that pitch raising may be construed in some languages as tone, and in others - such as the ones under current examination - as stress. Of course, such a type of analysis can be also maintained in a prominence-based account, thus on its own, it does not form an argument against prominence.

In the light of the above then, what exactly makes prominence an unsatisfactory account? While Hayes defines prominence as “perceptual salience” (1995: 271), this definition is undermined immediately afterwards when he says that: “The proposal here is to formalize this distinction (i.e. between weight and prominence), accounting for weight with a theory of quantity, based on moraic structure; and a theory of prominence, based on a different representation, which encompasses the whole set of phonetic properties (*weight included*<sup>4</sup>) that make syllables sound louder” (Hayes 1995: 271-272).

While it is obvious that the boundaries between stress and prominence are hazy, this statement seems to suggest that weight is a subtype of prominence and as a result, reference to weight and analyses based on it should be preferred since they are more restrictive. An added advantage in analysing data like the Pirahã ones by means of weight is that they now need not be distinguished from others where it has been robustly argued that syllable weight is involved. The phenomenon is basically the same, but what differs is its exact manifestation, e.g. onset vs. coda weight.

Moreover, weight makes more specific predictions than prominence. Given that prominence is such a broad notion under the tag of which a large number of phenomena may be encompassed, we could expect that it can easily over-generate. This is indeed what happens. For instance, under prominence it should be possible to find a language whose syllables with an onset *s* attract stress more than others as a result of *s*’s loudness. Since ‘louder’ can easily entail ‘more prominent’, this is a predictable system, and yet - to my knowledge - unattested. No similar expectation arises in a weight-based account.

---

<sup>4</sup> Emphasis added is mine.

In addition, weight is more restrictive in another way too. It can maximally reach two or three moras, but there is no similar upper bound for prominence. Virtually anything goes, therefore a considerable over-generation of patterns is predicted.

Although till now we focused on the ways prominence over-generates, it is time to draw on perhaps the most important argument against prominence, that of under-generation. Despite the negative consequences prominence bears, it is still possible to account for onset-sensitive stress by means of prominence as it has happened in most analyses including more recently de Lacy (2000) and Smith (2005). The problem appears once we look outside the domain of stress and examine clearly weight based phenomena like the ones discussed in §1.1, e.g. word-minimality, reduplication, etc. One can see that for these phenomena too, onsets may be relevant. These phenomena mainly include word-minimality, compensatory lengthening, geminates and less robustly metrics and reduplication. For such cases, as we have previously explained, only weight can be implemented, thus rendering any prominence analysis unsuitable and unable to account for the relevant facts.

It is thus a crucial part of this thesis to discuss in detail data that illustrate such cases since these not only corroborate the onset moraicity proposal but also receive more satisfactory accounts when compared to previous analyses that lacked the onset weight component. To this end, I will refer to Bella Coola word minimality (Chapter 4; Bagemihl 1991, 1998), Samothraki Greek compensatory lengthening (Chapter 5; Hayes 1989, Katsanis 1996, Kavitskaya 2002) and geminates in languages such as Pattani Malay (Chapter 6; Hajek and Goedemans 2003), Trukese (Chapter 6; Curtis 2003) and Marshallese (Chapter 6; Abo et al. 1976, Hendricks 1999), while more speculative data are discussed in Chapter 7 with respect to metrics, e.g. Luganda (Fabb 1997), reduplication and geminates e.g. Bellonese (Elbert 1988), as well as some additional cases of compensatory lengthening and word minimality. Some representative data for each of these cases are offered here. I will start with minimality in Bella Coola.

(14) *Bella Coola Word Minimality*

Word Shape	Examples	
a. V	*	
b. VV	ya	“good”
c. VC	ɲʌʔ	“dark, night”
d. CV	ʌʔi	“fast”

Bella Coola minimal words are bimoraic, which straightforwardly accounts for the well-formedness of (14b-c) and the unacceptability of (14a). The inclusion of (14d) among acceptable minimal words is peculiar, unless this is bimoraic too, i.e. the onset is moraic.

(15) *Samothraki Greek CL after onset r loss*

- a. singleton r-deletion word-initially  
rafts > á:fts 'tailor (masc.)'  
riyaŋ > í:yaŋ 'oregano'
- b. r-deletion in a complex onset cluster  
prótos > pó:tus 'first'  
ðédro > ðédu: 'tree'

In Samothraki Greek onset *r* deletes in certain positions and CL occurs, causing lengthening of the following vowel. This is compatible with the assumption that *r* carries a mora, which is preserved through lengthening when the /r/ itself is lost.

(16) *Trukese initial geminates and minimal words*

	<u>Form</u>	<u>Gloss</u>
a. CVV words	maa	'behaviour'
	oo	'omen'
b. CCV words	tto	'clam sp.'
	kka	'taro sp.'

In Trukese, the minimal word again is bimoraic. It can either take the form of CVV or CCV words, where the latter include a geminate. Geminates must then contribute a mora, and since they are word-initial, it is possible to analyse them as moraic onsets.

(17) *Pattani Malay initial geminates and stress*

- a. buwóh 'fruit'  
jalé 'street/path'
- b. b:úwóh 'to bear fruit'  
j:ále 'to walk'

Pattani Malay too has initial geminates as shown in (17b). The minimal pairs in (17a) and (17b) reveal that geminates have to be weight-contributing since they attract stress (17b). Again an analysis that represents initial geminates as moraic onsets seems in order.

As for (some of) the remaining cases, in the Ralik dialect of Marshallese, gemination is one of the mechanisms used to express the distributive form as in *korap* > *yokkoraprap* ‘gecko’, *diylah* > *yiddiylahlah* ‘nail’, *nib* > *yinnibnib* ‘preemptive’. Such reduplication could be merely seen as the addition of a mora. There is also evidence from stress which supports the idea that geminates are wholly syllabified in the onset and are weight-contributing. In other words, they are moraic onsets. Similar data occur in Bellonese too. Finally, in Luganda court songs, short vowels preceded by complex onsets whose second consonant is a glide seem to count as two moras for poetic meter purposes. We will explore these data extensively in subsequent chapters.

For the time being, what I have attempted to show in this section is that prominence accounts face certain difficulties and present numerous disadvantages. An exclusively weight-based proposal fares better as it is more restrictive and has a broader coverage of phenomena. On the contrary, prominence is simultaneously far too general - since it allows for lots of unattested patterns - and far too restricted by virtue of its intrinsic inability to account for weight-based phenomena.

Next, I will summarize Smith (2005), perhaps the most compelling prominence-based account, and as such, worthwhile of more detailed discussion. I will show that despite its advantages, it falls short of the range of onset-sensitivity effects. In §1.3.3, I will explore another approach, that of Gordon (2005), that makes use of both prominence and weight to account for onset-sensitive stress. Although the author attributes phonetic prominence properties to certain types of onsets, he then builds a phonological analysis based on featurally specified timing units. I will briefly review this approach and highlight some general problems it carries along. More detailed discussion, in particular with regard to technical issues, is deferred until chapter §2.2.3.

### **1.3.2 Smith (2005)**

Smith’s (2005) work originates in research that distinguishes positions into ‘strong’ and ‘weak’ (Casali 1996, Beckman 1999 among others). Strong positions are those which preserve contrasts, whereas in weak positions, neutralization is found, a phenomenon termed ‘positional neutralization’. For instance, in English or Catalan stressed syllables (strong positions), the full vowel inventory is found, but in unstressed syllables (weak positions), only a subset of it emerges. Strong positions are further grouped into two categories: i) phonetically strong positions containing stressed syllables (σ), onsets and

long vowels, and ii) psycholinguistically strong positions including roots and initial syllables ( $\sigma_1$ ).

Smith only focuses on strong positions and claims that strong positions need to be augmented, i.e. acquire properties that enhance their perceptual salience. In other words, prominent positions need to become more prominent, a phenomenon dubbed 'positional augmentation'. Properties that enhance prominence are the presence of: weight (HEAVY $\sigma$ ), stress (HAVE STRESS), high tone (HTONE), [Place] features in consonants (HAVECPLACE), onsets (ONSET) or onsets of low sonority (\*ONSET/X), and nuclei of high sonority (\*PEAK/X)<sup>5</sup>. Cross-classifying these properties with the strong positions above generates a number of constraints relativized to the position under consideration. For instance, ONSET/ $\sigma$ , \*[ONSET/X]/ $\sigma$ , HTONE/ $\sigma$ , requires that a stressed syllable has an onset, a low-sonority onset or high tone respectively.

However, not all possible combinations of prominent properties relativized to strong positions are available. Smith (2005) argues that only the constraints which pass the two filters she introduces are available. These filters impose limitations on the formal constraints CON may generate and stem from considerations regarding articulation, perception and processing, i.e. domains external to formal phonology (2005: 12). The first filter is the *Prominence Condition* which states that only general markedness constraints which increase prominence can be used to build markedness constraints specific to strong positions. The second filter is the *Segmental Contrast Condition* and applies to psycholinguistically strong positions. It bans augmentation constraints which alter properties that are critical for the identification of segmental contrasts during early-stage word recognition, thus a constraint like [\*PEAK/X]/ROOT demanding high-sonority syllable nuclei in roots is prohibited. However, the *Segmental Contrast Condition* allows augmentation constraints of a prosodic nature since these are not crucial during early-stage word recognition (e.g. HAVESTRESS/ROOT is allowed) or allows constraints that demarcate the left edge of the word, such as ONSET/ $\sigma_1$ , which requires that word-initial syllables are onsetful.

Smith's work is beyond doubt an important contribution to the field, with well worked-out predictions and good empirical coverage. However, it rests on certain assumptions whose validity is debatable or carries certain shortcomings which could

---

<sup>5</sup> The 'X' in these constraints stands for any step on the segmental sonority hierarchy. The version of sonority Smith assumes is in (19). Notably the \*PEAK/X hierarchy takes the form of \*PEAK/voiceless obstruents >> ... >> \*PEAK/low V, thus preferring segments of the highest sonority as nuclei. The \*ONSET/X hierarchy is the reverse, i.e. \*ONSET/low V >> ... >> \*ONSET/voiceless obstruents, favouring voiceless obstruents as best possible onsets (cf. Prince and Smolensky 1993/2004).



endanger the enterprise. Given that this model relies on formal constraints and filters whose sources are considerations external to formal phonology, it proposes an amalgamation of functional explanations and formal representations. This on its own is not necessarily disadvantageous - in fact it could be very much appropriate - however, it seems that the filters are essentially imposed in order that they stop certain - undesirable - constraints like [*\*PEAK/X*]/ROOT above from being generated.

In addition, within the filters themselves, extra statements are introduced to provide the conditions under which a constraint can nonetheless escape the violation of a filter. For instance, although the well-attested ONSET/ $\sigma_1$  constraint would normally fail the *Segmental Contrast Condition*, since it actually alters the segmental content of  $\sigma_1$ , it is nonetheless admitted, because of a stipulation that constraints demarcating the left edge of the word are not penalised by that filter. Although the existence of such a condition is justified (by means of facilitating word-boundary recognition through the augmentation of initial syllables in terms of onsets), this does not seem much else than an *ad hoc* move to permit a constraint that would otherwise be inadmissible.

Moreover, if it is true that the formal system is limited by functional considerations of external factors in terms of filters, then is it possible that choosing different considerations to implement as filters could lead to the generation of a very different set of constraints and consequently a different set of predicted linguistic patterns? Thus, while Smith (2005) may very well be right in the filters she uses, I contend that their designation as the suitable ones must be further justified.

As Blumenfeld (2005a) notes, an additional problem arises with Smith's neurophysiological definition of prominence, which picks out a stimulus as more salient than another if it "elicits a neural response of greater magnitude" (Smith 2005: 45). While this seems rather specific in some ways, it is vague in others (Maira Yip, p.c.), since the idea of "magnitude" can be variably interpreted, as e.g. a faster or more intense response or a response that causes activity in a larger area in the brain. "Magnitude" thus requires a more adequate definition. For this reason, evaluation particularly with respect to the *Prominence Condition* is not always straightforward.

Moving on to the discussion which is more central to the topic of this thesis, Smith discusses two constraints relating to onsets, namely: ONSET and *\*ONSET/X*, both of which can be relativized to initial and stressed syllables. Thus we get:

- (18) a) ONSET: Syllables have onsets  
           ONSET/ $\sigma_1$ : Initial syllables have onsets  
           ONSET/ $\acute{\sigma}$ : Stressed syllables have onsets  
       b) \*ONSET/X: Syllables have less sonorous onsets  
           \*[ONSET/X]/ $\sigma_1$ : Initial syllables have less sonorous onsets  
           \*[ONSET/X]/ $\acute{\sigma}$ : Stressed syllables have less sonorous onsets

For ONSET and \*[ONSET/X], Smith claims that there is neurophysiological evidence suggesting that the presence of an onset and its specific quality enhance prominence<sup>6</sup>. Her argument is roughly the same as in Gordon (2005) who makes reference to ‘adaptation’ and ‘recovery’, an idea that runs into problems, as I will show in the next section in detail and thus will not discuss here to a larger extent.

Despite neurophysiological support on their existence, some of these constraints do not receive much support empirically, especially the version of \*[ONSET/X] pertinent to stressed syllables, i.e. \*[ONSET/X]/ $\acute{\sigma}$ . With this constraint, Smith generates the patterns of Pirahã where low-sonority onsets attract stress, as well as of Niuafu`ou (de Lacy 2000) where glide onsets are avoided in stressed syllables.

Notably, these patterns make reference to the *extreme* edges of the sonority hierarchy (19), but if Smith were right, and sonority alone is the reason one finds the Pirahã and Niuafu`ou patterns, then we should be able to find languages which utilize any of the *intermediate* cut-off points for the \*[ONSET/X]/ $\acute{\sigma}$  constraints as well.

(19) *Sonority Hierarchy with respect to consonants (Smith 2005: 56)*

glides > rhotics > laterals > nasals > voiced obstruents > voiceless obstruents

For instance, in analogy to Niuafu`ou, we would predict that there is a language where glides, rhotics and laterals are avoided in stressed syllables or another where all sonorants do, and so on. Similarly, a pattern reminiscent of Pirahã could arise where voiceless obstruents, voiced obstruents and nasals attract stress to the exclusion of all other consonants. As far as I know, none of these arises, which suggests that the onset sensitivity effect is not down to sonority per se, as Smith (2005) argues. In the next section, I will claim that the effect should be attributed to voicing considerations (i.e. a small part of the sonority hierarchy), therefore the only pattern we should get is where voiceless onsets attract stress more than voiced ones. Data from Pirahã, Arabela and Karo suggest that this is on the right track.

---

<sup>6</sup> In fact, these are the only constraints for which explicit neurophysiological or psycholinguistic justification on their prominence is offered.

However, this still leaves us with Niuafu`ou, which is not predicted by a ‘voicing’ theory. I suspect that actually Smith herself (2003) offers a possible explanation. Smith makes a distinction between ‘true onset’ glides and ‘nuclear’ onglides. Simplifying a bit, the former are really in an onset position, whereas the latter are actually underneath the nucleus node. As a result, only the former are subject to constraints that make reference to onset structure (\*ONSET/X), whereas the latter escape them. It could thus be conceivable to build an analysis which treats Niuafu`ou glides as nuclear onglides. As such, they would be subject to a different set of constraints, i.e. probably the [\*PEAK/X]/σ constraints, and consequently would not interfere with the genuine onset-sensitive stress effects.

All in all then, Smith’s (2005) proposal, although largely successful, poses some conceptual issues and faces some difficulties with respect to the onset-sensitivity constraints it imposes that render it inadequate to deal with the phenomenon at hand. And of course, as with all prominence accounts, it also suffers from the inability to deal with clearly weight-based phenomena that involve onsets.

### 1.3.3 Gordon (2005)

Gordon (2005) examines the phenomenon of onset sensitivity from its phonetic perspective making use of the notions of adaptation and recovery (Delgutte 1982, Viemeister 1980). The term “adaptation” expresses the decline in auditory sensitivity to a stimulus while “recovery” expresses the fact that sensitivity to a new stimulus can be regained by means of a period of silence or reduced acoustic energy. Onsets - and in particular less intense and thus less sonorous onsets - provide exactly this recovery phase best and enhance the perceptual energy of the rime. Thus, (low sonority) onsets can contribute to weight through the perceptual boost they offer to the following vowel.

Gordon’s approach, although appealing, presents a number of problems. First, it is not clear why adaptation and recovery should necessarily refer to onsets. All the system needs to be interested in is regaining sensitivity to a new stimulus, but that could equally be provided by a coda. In fact, syllable position seems irrelevant. Second, the correlation between onset weight and the auditory boost offered by onsets does not have to be a necessary one, although it is presented as such. By saying that “Given the auditory boost provided by an onset consonant, it is not surprising that syllables containing an onset consonant might be phonologically *heavier* than onsetless syllables in certain languages” (2005: 604), Gordon seems to suggest that this is an inescapable

conclusion and does not argue further for it. But it is not clear why this property does not merely increase the prominence of a syllable that contains (such) an onset, and instead has to be construed as weight. In fact, Gordon's next sentence seems to suggest the same, as he explains that less intense onsets produce higher auditory boost and thus are more likely to make their syllable *more prominent*. Such conflation of weight and prominence is actually a recurring theme, as we will also see in more detail (cf. §2.2.3) when we examine Gordon's constraints that account for onset weight.

One important proposal is that rimal weight takes precedence over onset weight cross-linguistically (2005: 605). This is because the effect of the onset on the perceived loudness of a sound due to the mechanisms of adaptation and recovery is really minor compared to the rime material. Given that a long vowel or a coda consonant can increase auditory prominence to a much larger extent than a low sonority onset, rimal weight is prioritized. This is problematic for three reasons.

First, in Arrernte<sup>7</sup>, it is actually onset weight which according to Gordon has primacy. He attributes this reversal to certain phonetic properties of individual languages which impede rimal weight from taking priority, e.g. in Arrernte onsetless vowels are shorter than onsetful vowels and vowel lengthening occurs in open syllables with an onset. At any rate, the generalisation above is weakened by such facts.

Second, not much discussion of the auditory boost a coda could offer is presented, although it is claimed that its presence significantly increases auditory prominence (p. 605). One thing that is mentioned though is that "greater sonority in the rime helps to offset the reduction in perceptual energy due to adaptation" (p. 633). It does not seem clear why high-sonority should have this effect in that position. Shouldn't we again need low-sonority codas to achieve recovery? A more general consideration however is that if a coda is phonetically more effective than an onset, then why should there be onset weight effects in the first place?

Third, Gordon's claim about the primacy of rimal weight is not really accurate. As I hope to show, many of the languages he discusses as involving onset weight (e.g. Aranda, Alyawarra, Bislama, Nankina and others) do not really show this convincingly. As for the rest, e.g. Pirahã, it is true that rimal weight takes precedence over onset weight, but we also need to take into account that Pirahã lacks codas altogether. Therefore, rimal weight really translates to nucleic weight<sup>8</sup>. It is well-known that in the

---

<sup>7</sup> The relationship of Arrernte and Aranda is not clear (cf. discussion in the Ethnologue: <http://www.ethnologue.org/>), but are usually referred as alternate names of the same language.

<sup>8</sup> According to my proposal, Arabela and Karo are the other truly onset-sensitive stress languages, which nonetheless Gordon does not discuss. The former seems to lack codas, although data are not absolutely

more common case of languages where onsets contribute no weight, there are some where nonetheless nucleic weight takes primacy over short nucleus+coda weight, i.e. VV > VC as in Klamath, Chickasaw (Gordon 2004), and Kashmiri (Morén 1999/2001, 2000). This is then not a result that should be considered intrinsic to onset-weight stress languages, as Gordon seems to suggest.

As we have seen, onsets are claimed to enhance the perceptual energy of the rime through the recovery period they offer. Onsets however themselves do not contribute to the auditory prominence of the syllable, but only through their effect to the following rime. This is a “crucial feature” (Gordon 2005: 604) of this analysis, which is never justified. So the perceptual energy of the onset itself is not addressed, but it is likely that it too could have effects on stress. In particular, perceptual energy (with respect to the rime) “is quantified as the summation of loudness over time: a rime’s energy is thus a function of both duration and its intensity” (2005: 606). If the same definition applies to the perceptual energy of the onset itself, then we would expect it would be enhanced when sonorants or fricatives appear as onsets, but not if stops appear as onsets.

Some further expectations with respect to perceptual energy are presented below. Suppose we consider a couple of CV sequences and for each we examine three windows. Window 1 will be the perceptual energy of the onset itself, window 2 refers to the recovery period, whereas window 3 is the steady state of the vowels, which can be disregarded since it is expected to be the same in all cases. As explained above, in window 1 the perceptual energy will be increased when the onset is made up of e.g. a nasal compared to e.g. a voiceless stop. On the other hand, low sonority onsets such as voiceless stops will contribute to recovery (window 2) more significantly than more sonorous onsets such as nasals will.

(20)

	<b>Window 1</b> <i>perceptual energy of the onset itself</i>	<b>Window 2</b> <i>recovery</i>	<b>Window 3</b> <i>steady state of the vowel</i>
NV	+	-	
PV	-	+	

N=nasals, P=voiceless stops

*For Window 1:*      + = high perceptual energy  
                             - = low perceptual energy

*For Window 2:*      + = high recovery  
                             - = low recovery

---

clear in this respect (cf. Rich 1963), whereas in Karo I will claim that only sonorant codas are moraic (§2.4.3.2.2).

The problem that Gordon faces is that he does not take into account window 1. In his system it seems to be the case that the effects during the recovery period (window 2) take priority over these of the onset perceptual energy (window 1) predicting that voiceless stop onsets make better weight-bearers than nasals, i.e.  $PV > NV$ . However, it could be possible that in other languages window 1 matters more than 2 resulting in a situation exactly opposite from the one discussed above. In this case more intense, thus more sonorous onsets, like nasals, would be predicted as more able to carry weight,  $NV > PV$ . This issue is left unexplored. Note that if one also considers similar simultaneous effects from codas, very complex and presumably unlikely systems are produced.

The link between onset sensitivity and perceptual energy, which is a key factor in Gordon's approach, generates additional questions regarding the predictions made. For instance, does he make the prediction that in languages lacking onset sensitivity, perceptual energy should not be of particular importance? Should we also anticipate that languages where perceptual energy is of significance should also be onset-sensitive? It is unclear what the expectations should be, although Gordon discusses languages whose stress is quantity insensitive and which consistently prefer rimal weight. This can only be expected if this is phonetically superior to onset weight. This may very well be true, although I have some reservations over the use of rimal weight and the equation of (long) nucleic weight with VC weight.

Other claims, which are explicitly stated, seem quite dubious too. For example, Gordon proposes that less intense onsets should provide a bigger auditory boost. Within this class of onsets, he also includes obstruents, but although stops indeed have low intensity, the same does not hold for fricatives which are rather intense sounds. In this view it would be surprising to find languages where fricatives pattern with the weightful onsets and not the weightless ones. However, in Pirahã, *s* patterns with the voiceless stops and contributes weight to the syllable.

For all these reasons (but also see §2.2.3), Gordon's approach seems inadequate to deal with onset effects satisfactorily. Having dispensed both with pure prominence and mixed prominence-weight accounts, I will now move on to the third possibility, that of onset weight, which is the one I will entertain in this thesis.

## 1.4 Onset moraicity

### 1.4.1 Types of moraic onsets: distinctive vs. coerced weight

Suppose then that we do accept the possibility that onsets may be moraic. How exactly is this implemented? What kind of moraic onsets are there? I would like to propose that there are two types of moraic onsets. This classification resembles Morén's (1999/2001) categorization between distinctive and coerced weight. The former describes the situation where there are phonemic weight distinctions, while the latter refers to weight acquired in the output as a result of a requirement such as word minimality, weight-by-position etc. Although Morén only examines coda weight<sup>9</sup>, since he assigns no weight on onsets, I would like to suggest that a similar categorization is applicable for the case of onsets too.

Thus both distinctive and coerced moraic onsets occur. The former are in fact geminates, but rather than being syllabified in a coda-onset configuration, they are instead wholly syllabified in the onset. I will return to this point in §1.4.1.3 and provide some justification for such syllabification. The second type of moraic onsets are those of coerced weight. In this case there is no evidence for an underlying contrast between singletons and geminates, but on the surface some onsets do appear moraic.

Distinctive weight refers to an underlying weight contrast. For instance, the contrast between Hungarian *vice* 'janitor' vs. *vic.ce* 'his joke' is one that in the input will be represented as /vice/ vs. /vic<sup>h</sup>e/ respectively. In coerced weight however, there is no such contrast. Instead, consonants will be found to be moraic on the surface due to a high imperative such as Word Minimality, Stress-to-Weight, or Weight-by-Position etc. Morén argues that this difference on the nature of weight also has repercussions on what kinds of moraic consonants can be found. In distinctive weight there are no expectations as to which type of segments will show contrasts between short versus long. Since weight is lexically specified in the input, this will be unpredictable and consequently perhaps arbitrary.

But the same principle does not hold in coerced weight, which is moulded on the surface where markedness constraints are applicable. There, sonority considerations become relevant. Hierarchies referring to *nucleus* and *coda* moraicity show that more sonorous segments are preferred to be moraic as well as attract stress, leading to the

---

<sup>9</sup> His discussion also applies to vowels, but as I am only concerned with consonants, I set it aside.

conclusion that the more sonorous a coda segment is, the more likely it is for it to be moraic (Zec 1988, 1995, Morén 1999/2001)<sup>10</sup>.

Thus, according to Morén, the famous example of Kwakwaka (Zec 1988, 1995) where the moraicity of codas in CVC syllables is variable depending on whether the codas are sonorous, is an instance of coerced weight. In particular, non-glottalised sonorant codas are moraic, whereas glottalised sonorants and obstruents are not. The prediction then is that in languages with coerced coda weight, sonority is crucial and thus it should not be possible for non-sonorous codas to be moraic with more sonorous ones being non-moraic.

However, as we have seen, in languages with distinctive weight, no similar restrictions apply. Lexical specification occurs and therefore sonority reversals with respect to moraicity are admitted. For instance, in languages that allow only one type of geminates, these are commonly the less sonorous voiceless rather than the more sonorous voiced ones, e.g. Lak, Nez Perce, Ojibwa, Totonac, Yakut and Japanese (Morén 1999/2001).

If this idea is on the right track, then we would expect similar effects in onset moraicity too. Languages like Trukese, Pattani Malay and perhaps Marshallese and Bellonese have geminates that can be analysed as *distinctive* moraic onsets and therefore we should have no expectations with regard to their sonority. In most of these cases, all consonants are able to appear as singletons or geminates with the possible exception of Pattani Malay, where a subset of the consonants cannot geminate. But this is exactly what is anticipated if geminates are lexically specified as such. These languages will be dealt with in more detail in Chapter 6 (Pattani Malay, Trukese and Marshallese) and Chapter 7 (Bellonese).

As for *coerced* moraic onsets, these emerge in Pirahã, Arabela and Karo (Chapter 2) due to MORAIC ONSET (39), the equivalent of Weight-by-Position for onsets, and in Bella Coola due to Word Minimality (Chapter 4). In these cases, we will expect that onsets will follow specific sonority considerations (with respect to voicing, as I will show immediately below). The sonority profile of coerced moraic onsets is explored in the next section and exemplified in Chapter 2 in the discussion of the effects of moraic onsets on stress. With respect to moraic onsets in Bella Coola, we will see that all consonants can serve as moraic onsets, a fact that is compatible with the coerced

---

<sup>10</sup> In the next section, I will argue that this conclusion is right when we examine nucleus and coda weight (rimal weight). For onset weight, the opposite obtains, which is why I will relate moraicity not with sonority per se, but with syllable well-formedness. For details, see below.



weight expectations. In summary then, the useful distinction between distinctive and coerced weight seems applicable in onset weight too.

#### 1.4.1.1 *The effect of voicing on tone and stress and its consequences for onset moraicity*

In the previous section, we have seen that the preferable moraic codas are the ones which are more sonorous. This is by no means a coincidence, since it relates to the issue of what optimal nuclei and margins look like. Prince and Smolensky (1993/2004) argue that optimal peaks are high in sonority, whereas optimal margins are low in sonority. This is the result of the harmonic peak and margin hierarchies in (21).

- (21)     $\text{Peak}_{\text{syll}}$              $a > e, o > i, u > \dots > p, t, k$   
           $\text{Margin}_{\text{syll}}$          $p, t, k > \dots > i, u > e, o > a$

Technically, these can be translated into the following constraints (for the harmonic alignment schema see Prince and Smolensky 1993: 136 and McCarthy 2002: 20-22):

- (22)    Peak prominence:  $*\text{PEAK}/p, t, k >> \dots >> * \text{PEAK}/i, u >> * \text{PEAK}/e, o >> * \text{PEAK}/a$   
          Margin prominence:  $*\text{MAR}/a >> * \text{MAR}/e, o >> * \text{MAR}/i, u >> \dots >> * \text{MAR}/p, t, k$

Note though that these hierarchies and fixed rankings are not exactly accurate; as we have seen, codas actually prefer to be quite sonorous and at any rate much more sonorous than optimal onsets (Murray and Vennemann 1983, Vennemann 1988, Zec 1988, 1995, Clements 1990, Smith 2003). Bearing this in mind, the above hierarchy needs to be refined as below (see Baertsch and Davis 2003 for a similar view)<sup>11</sup>:

- (23)    *Syllable-constituent well-formedness*  
          Nucleus prominence:  $*\text{PEAK}/p, t, k >> \dots >> * \text{PEAK}/i, u >> * \text{PEAK}/e, o >> * \text{PEAK}/a$   
          Coda prominence:  $*\text{CODA}/p, t, k >> \dots >> * \text{CODA}/i, u >> * \text{CODA}/e, o >> * \text{CODA}/a$   
          Onset prominence:  $*\text{ONS}/a >> * \text{ONS}/e, o >> * \text{ONS}/i, u >> \dots >> * \text{ONS}/p, t, k$

<sup>11</sup> Of course the combined interaction of syllabification considerations, the requirements on what can constitute an onset or a coda as shown in (23) and the high-ranked constraint NUC which demands that syllables have nuclei, ensure that the more sonorous segments syllabify as nuclei and not as codas (e.g. usually only high vowels can be glides and syllabified in margins, but nothing more sonorous than that). In other words, although low vowels are both the best possible nuclei and codas according to (23), we do not predict that low vowels will be syllabified in a coda. This is simply because the coda and peak hierarchies interact with one another and the requirement to have a nucleus will give priority in picking out the low vowel as a nucleus and not as a coda (for a similar argument, cf. Baertsch and Davis 2003).

This correctly generates the fact that optimal nuclei and codas are high in sonority, whereas optimal onsets are low in sonority. Moreover, observe that preferable moraic codas, and - of course - nuclei, are the ones which are high in sonority (Zec 1988, 1995, Yip 1992). One would perhaps expect then that the more sonorous a segment is, the more likely it is to be moraic. While this works fine for codas and nuclei, it does not account equally well for the newly introduced moraic onsets. The problem is that in onsets, as the empirical data suggest, less sonorous segments (i.e. the voiceless segments) are moraic, despite being less sonorous than their voiced counterparts. But this point only proves problematic if one makes a direct connection between moraicity and sonority. If one instead relates moraicity with sonority indirectly via syllable well-formedness, then facts receive a natural explanation.

What I mean by that is that if we allow ‘more well-formed’ to entail ‘more likely to be moraic’, then in codas, we correctly expect that sonorous codas are the ones which tend to be moraic like in Kwakwala, since for codas (and nuclei), more well-formed implies more sonorous. But the reverse holds for onsets, where least sonorous segments make better onsets. The expectation in that case is that less sonorous onsets will be more likely to be moraic by virtue of their well-formedness.

In this view, moraicity is concomitant with syllable well-formedness and not with sonority per se. Sonority is decisive in terms of syllable well-formedness, but is also sensitive to sub-syllabic constituenthood. Depending on the syllable constituent, sonority regulates well-formedness differently, i.e. low-sonority = well-formed in onsets; high sonority = well-formed in codas. Consequently, high sonority segments make the best moraic nuclei and codas, while low sonority segments make the best moraic onsets. This can be schematized in the following hierarchy:

(24) *Moraicity as a function of well-formedness*<sup>12</sup>:

Nucleus moraicity: \*μ/PEAK/p,t,k >> ... >> \*μ/PEAK/e,o >> \*μ/PEAK/a

Coda moraicity: \*μ/CODA/p,t,k >> ... >> \*μ/CODA/e,o >> \*μ/CODA/a

Onset moraicity: \*μ/ONS/a >> \*μ/ONS/e,o >> ... >> \*μ/ONS/p,t,k

Given that our focus here is on margins, the part of the moraic markedness hierarchy we are interested in can be informally summarized as follows.

---

<sup>12</sup> Three-predicate constraints have been used in other works too, e.g. de Lacy (2000).

(25) *Moraic markedness for margins - first approximation*

Coda moraicity: \*μ/CODA/non-sonorous >> \*μ/CODA/sonorous

Onset moraicity: \*μ/ONS/sonorous >> \*μ/ONS/non-sonorous

As I hinted earlier however, there is one particular dimension of sonority which is relevant for onset moraicity purposes, namely voicing. Manner features for instance seem inert. As a matter of fact, I would like to propose that the only part of the sonority scale relevant for onset moraicity purposes is voicing. An appealing explanation is available for this fact.

More specifically, it is well known that voicing becomes relevant in many tone languages, so that voiceless onsets usually raise the pitch of the following vowel, whereas voiced ones lower it. In voiceless obstruents, vocal folds are tenser, a fact that is linked with pitch raising. The reverse occurs in voiced obstruents where slacker vocal folds lead to pitch lowering (on the phonetics of this see Yip 2002: 6-7 and references cited therein). In fact, depression of  $F_0$  after voiced stops is very likely universal as Kingston and Solnit (1988b) state.

During the course of history, numerous languages have lost the voicing contrast and have re-interpreted it in terms of tones. A common scenario is that such a language now only has voiceless onsets and a contrast between L and H tones. L tones emerge where there used to be a preceding voiced onset, whereas H tones where there was originally too a voiceless onset. In fact, this ‘evolutionary’ path can be synchronically traced among dialects of the same language. Yip (2002: 35, citing Svantesson 1983) notes that in the southern dialects of Kammu there is still a voicing contrast in onset position for the obstruents, but this no longer exists in the northern dialects. There, a H tone appears in positions that correspond to voiceless obstruents and a L tone in positions that correspond to voiced obstruents in the southern dialects. This pattern is illustrated below.

(26) *Kammu dialects*

<u>South</u> (voicing contrast-no tone)	<u>North</u> (no voicing contrast-tone contrast)
klaaŋ	kláaŋ ‘eagle’
glaaŋ	klàaŋ ‘stone’

Other languages still maintain the voicing contrast, but consistently show lower pitch after a voiced obstruent and higher pitch after a voiceless obstruent. One example of this type is in Songjiang, a Wu dialect of Chinese (examples again from Yip 2002: 7). 5

denotes the highest pitch, 1 is the lowest one. The first digit marks the pitch at the beginning of the syllable, the second marks the one at the end of the syllable.

(27) *Songjiang tones*

ti	53	'low'	di	31	'lift'
ti	44	'bottom'	di	22	'younger brother'
ti	35	'emperor'	di	13	'field'

So far, we have seen that voicelessness goes hand-in-hand with H tone whereas voicing is concomitant with L tone. An additional important point is that in some languages H-toned moras attract stress (de Lacy 1999, 2002). One example of this sort is Golin, where stress falls on the rightmost H-toned syllable. In the absence of H-toned syllables stress is word final.

(28) *Golin H-tones and stress* (de Lacy 1999; stress marked in bold)

LLL	kàw.lì.gì	'post'
HLL	á.kò.là	'wild fig tree'
LHL	gò.má.gì	'type of sweet potato'
LLH	ò.nì.bá	'snake'
HHL	sí.bá.gì	'sweet potato type'
HLH	én.dè.rín	'fire'

In the light of the above, the idea proposed here then is rather simple; while some languages make use of the pitch perturbation caused by voicing to convey tone, some others use it to convey stress. The fact that in the latter cases, stress is involved rather than tone is supported by the fact that some of these languages also have tone, e.g. Karo (29) and Pirahã (30) or present groupings into structures that look like feet, as in Arabela (31).

In Karo there is a phonological contrast between high and low tones. Low tones are phonetically realized as mid when stressed. Some minimal pairs are presented [acute accent marks H tone; lack of accent indicates phonological L tone (but phonetic mid)].

(29) *Karo contrastive tones* (Gabas 1999)

a.	cá <sup>d</sup> n	H-tone	'to wash'	vs.
	ca <sup>d</sup> n	L-tone	'to pluck'	
b.	tóy	H-tone	'to disappear'	vs.
	tóy	L-tone	'to see'	
c.	pén	H-tone	'to open'	vs.
	pén	L-tone	'to step'	

Pirahã additionally provides evidence that tone is independent of stress. To show that, we merely need to consider pairs of disyllables of the type PVBV and PVPV (P=voiceless, B=voiced), where the only thing that changes is the tone. Stress is not affected. In PVBV, the first syllable is stressed by virtue of the stress-attracting voiceless consonant, whereas in PVPV, both syllables are of the same type, therefore the rightmost one is stressed. The following cases illustrate (stress is denoted by boldface; accents mark tone).

(30) *Pirahã: tone independent of stress* (K. Everett 1998)

- i) **PVBV**
  - a. **HH:** tígí 'small parrot'
  - LL:** pígi 'swift'
  - b. **LH:** sabí 'mean'
  - HL:** ?ábi 'to stay'
- ii) **PVPV**
  - a. **LH:** tí?í 'honey bee'
  - HL:** tíhi 'tobacco'

Evidently both H and L tones are possible after voiced and voiceless consonants, which suggests that tone is not dependent on the preceding consonant's voicing. Moreover, we see that stress stays constant, whereas the tonal patterns may change, which clearly suggests that stress and tone are independent from each other.

In Arabela, there is no tone, but clearly there are stress groupings, given that secondary and primary stresses are available<sup>13</sup>. Since these groupings are partly determined by onset voicing, this is also a case where the pitch perturbation of voicing is interpreted as stress, not as tone. Some data exemplify this.

(31) *Arabela feet* (Payne and Rich 1988)

- a. (tèna)(kári) 'afternoon'
- b. (sàma)(rú) 'spirit'
- c. (hùwa)(hàni)(yá) 'peaceful'
- d. (nòwa)(fí)(fáno) 'brightened'

The advantage of using (the lack of) voicing as a sign for onset moraicity is two-fold. First, when languages use it to mark stress, they simply employ an already available phonetic cue (by means of pitch raising/lowering) which just happens not to be

<sup>13</sup> It has been suggested that apart from metrical feet, tonal feet also exist in languages such as Hausa, Northern Mande, Isixhosa (Leben 2001) or Kera (Pearce 2005). Leben (2001) however notes that the characteristic of a metrical foot is that it contains one stress per foot, whereas constituency in tonal feet is not as clear. Arabela obviously belongs to the languages with metrical feet.

associated with tone. This is merely phonologised in such languages through the presence of moras on onsets.

In fact, there is recent phonetic evidence which shows that pitch may be exclusively used to convey stress. The study of the prosodic properties of (Standard Modern) Greek TV news reports as conducted by Papazachariou and Politis (2005) reveals that certain features distinguish them from everyday narratives. In particular, in news reports, three of the typical four characteristics of stress, namely intensity, duration and vowel quality are disregarded. It is solely the fourth cue, i.e. pitch, that plays active role in marking stress. Although this study does not examine any possible effect consonantal quality may bear, it is significant to note that this work makes a direct, and - more importantly - practically exclusive correlation between stress and pitch.

There is another advantage to the association between onset moraicity and the lack of voicing. Being voiceless coincides with being less sonorous. Although sonority scales are notoriously varied (see Parker 2002 for a good overview) and there have been suggestions that voicing should not even be included in them (cf. Clements 1990), there are numerous versions of the sonority hierarchy that include voicing in it. But even so, there is no consensus on whether there is further subdivision based on the continuancy or not of the consonants. In particular the following possibilities have all appeared in the literature (again Parker 2002: 68-69 is a valuable source of relevant references).

- (32) *Sonority and obstruent voicing* ( > : means more sonorous than)
- i)       voiced fricatives > voiceless fricatives > voiced stops > voiceless stops
  - ii)       voiced fricatives > voiced stops > voiceless fricatives > voiceless stops
  - iii)       voiced fricatives > voiceless fricatives = voiced stops > voiceless stops

While the position of voiced fricatives and voiceless stops at the top and bottom of the scale respectively is established, the status of voiceless fricatives and voiced stops is more contentious. Parker also notes that no adequate empirical evidence is available to support any of these scales. Imdlawn Tashlhiyt Berber (ITB) can be used to support (32i), while Kolokuma Ijo<sup>14</sup> (Williamson 1965) backs up (32ii). Parker himself

---

<sup>14</sup> In this language stems are made of maximally three consonants C<sub>1</sub>, C<sub>2</sub> and C<sub>3</sub>. The quality of C<sub>1</sub> affects the quality of C<sub>2</sub> and the same holds between C<sub>2</sub> and C<sub>3</sub>. Consonants belong to one of these three categories: a) strong, i.e. voiceless obstruents, b) medium, i.e. voiced obstruents and c) sonorants. If C<sub>1</sub> is strong, C<sub>2</sub> can belong to any of the above categories. If C<sub>1</sub> is medium, then C<sub>2</sub> is either medium or weak, and finally if C<sub>1</sub> is weak, then C<sub>2</sub> is weak too. Similarly for C<sub>2</sub> and C<sub>3</sub>. Thus we can get something like *o-kosi* 'old person' (strong-strong) and *kodmu* 'moon' (strong-medium-weak), but not *\*o-losi* (weak-strong)

concludes that instrumental data suggest that the default situation in languages is the one depicted in (32ii). Nonetheless, there are languages like ITB, which require the reverse relationship, therefore the relation between voiceless fricatives and voiced stops is not universally fixed.

The data from Arabela, Pirahã and Karo are in full accord with Parker's conclusion, in that the whole voiced obstruent series is more sonorous than the whole voiceless counterpart. If he is indeed right, then these languages merely present the default situation. Thus, in at least these three languages the voicing distinctions are conflated from the four of (32ii) into the two below.

- (33) *Sonority hierarchy relevant for the languages under current consideration*  
voiced > voiceless

As we have mentioned before, consonants in the onset position (as far as stress and weight are concerned) are not sensitive to the manner features. Again there seems to be a parallel with the voicing and tone relationship, since to my knowledge, manner is similarly irrelevant in tone<sup>15</sup>.

The current proposal, as we have seen before, attempts to link moraic markedness and syllable well-formedness by making use of sonority. More concretely, in the case of moraic onset markedness it is not the whole sonority hierarchy that is relevant, but only the voicing chunk (33), as this is the one that interacts with pitch perturbations. Since more well-formed for an onset entails less sonorous, i.e. voiceless, and since more well-formed implies more likely to be a moraic, it follows that the moraic onset markedness hierarchy is informally translated to the one below.

- (34) *Onset moraic markedness hierarchy*  
\*μ/ONS/voiced >> \*μ/ONS/voiceless

This directly mirrors the fact that voiceless onsets make better moraic onsets than the voiced ones by virtue of being more well-formed.

Up to now, I have focused on the status of obstruents and left out the sonorants out of the picture. The question then is whether these pattern with voiced segments in

---

or *\*dokmu* (medium-strong-weak). These facts fall out if the strength hierarchy follows (inverse) sonority considerations with voiced consonants being more sonorous than voiceless ones.

<sup>15</sup> Yip (2002: 33-34) and Kingston and Solnit (1988b: 14-15) also mention some effects on tone due to aspiration, glottalization or breathy voicing, but at any rate these comprise laryngeal properties and not manner ones.

being non-moraic onsets, or with voiceless ones in being moraic? Actually, both options seem to be available as the data suggest. In Pirahã and Arabela, sonorants pattern alongside the voiced obstruents, as they do not attract stress, but in Karo along with the voiceless ones, since they attract stress.

A possible explanation for this distribution goes along these lines. Sonorants are spontaneously voiced, so that, as Rice and Avery (1989) put it, sonorants always include an SV (Spontaneous Voice) feature, which is part of the defining make-up of sonorants. In some instances, sonorants and voiced obstruents pattern in the same way, which is why in such cases, they must have a feature that groups them together. Unlike Rice and Avery (1989) who use [SV] as the relevant feature, I will make the more traditional assumption that the shared feature is either the binary [ $\pm$ voice] or the monovalent [voice], which is obligatory in voiced obstruents, but optional in sonorants (Itô and Mester 1986, Lombardi 1994).

Use of the binary feature obviously captures the match in behaviour between voiced obstruents and sonorants in Pirahã and Arabela and can be easily formulated as:  $*\mu/\text{ONS}/[+ \text{voi}] \gg * \mu/\text{ONS}/[- \text{voi}]$ , but is problematic when one considers Karo. In this language, sonorants are equally stress-attracting as voiceless obstruents are, so one could be tempted to say that sonorants are actually [-voice]. Apart from this being at the very least an unappealing and unlikely move, it also makes the prediction that sonorants should act as voiceless obstruents do in other processes, a fact, which in the case of Karo cannot be confirmed.

If we assume instead that the monovalent [voice] is employed, then again we expect sonorants to pattern with voiced obstruents so long as the latter are specified as [voice] too (like in Arabela and Pirahã). But in languages like Karo where this is not the case, sonorants can be considered to lack the [voice] feature altogether and in that way indirectly be grouped alongside the voiceless obstruents, which lack this feature too. However, since there is no feature present, it is not necessarily the case that these two natural classes should actively participate in other processes in the same way. I contend that a monovalent [voice] feature is better equipped to handle the cases considered, which is why the onset moraic markedness hierarchy in (34) is now updated as shown below.

(35) *Onset Moraic markedness - revised*

$*\mu/\text{ONS}/[\text{voi}] \gg * \mu/\text{ONS}$



Before leaving this part of the proposal, there are two more interesting observations we can make. First, we have argued that sonorants in some onset moraicity cases pattern with the voiced obstruents (Pirahã, Arabela), whereas in others with the voiceless obstruents (Karo). This is identical to what happens with respect to the voicing of consonants and tone, as Yip (2002: 37) notes: “... the behaviour of voiced sonorants, which in some languages pattern with voiced obstruents but in others with voiceless ones”. In addition, Kingston and Solnit (1988a, 1988b) observe that since sonorants do not automatically perturb the  $F_0$  of adjacent vowels, there is no phonetic reason “...to expect that the fundamental frequency of the following vowel will be either elevated or depressed by the sonorant’s laryngeal articulation; instead the speaker may be free to choose either elevation or depression of fundamental frequency”.

This dual behaviour of sonorants is thus entirely anticipated if we are talking about the same phenomenon with different manifestations, i.e. ‘voicing effects on pitch’ as cues for either tone (established) or stress (proposed). More concrete examples with respect to tone can be found in Bradshaw’s (1999) discussion of depressor consonants. She claims that the consonant-tone interaction always involves tone and consonants marked with the feature [voice]. Voiced obstruents always have such feature, therefore they are systematically depressor consonants. Voiceless obstruents never are (and few apparent counterexamples can be re-analysed by some other means), whereas sonorants more commonly pattern with the voiceless obstruents and in a few occasions with the voiced obstruents. While the default situation for sonorants is to lack any [voice] specification, whenever they pattern alongside the voiced obstruents, they are marked as [voice].

Bradshaw lists numerous languages where voiceless obstruents and sonorants pattern together to the exclusion of voiced obstruents<sup>16</sup>. These include: Suma, Siswati, Yaka, Miya, Digo, Bade and others. For instance in Bade a H spreads to a L-toned mora if that mora precedes another H-toned mora. In effect, /HLH/ emerges as [HH!H]. Spreading occurs if the intervening consonant is a voiceless obstruent (36a) or a sonorant (36b), but not if it is a voiced obstruent (36c).

(36) *Bade H-spread -- sonorants and voiceless obstruents pattern together*

- |    |   |              |
|----|---|--------------|
| a. | nón kàtáw → nón ká!táw                  | ‘I returned’ |
| b. | nón làwáw → nón lá!wáw                  | ‘I ran’      |
| c. | nón gàfáw → nón gáfáw      * nón gá!fáw | ‘I caught’   |

<sup>16</sup> A collective summary of the languages discussed and the pattern they present can be found in Bradshaw (1999: 6).

Still in (fewer) other languages, the sonorants' behaviour moves to the other direction. In Ngizim or Nupe, sonorants pattern like the voiced obstruents and not their voiceless counterparts. As an example, in Nupe the L tone of a nominal prefix or tense/aspect marker spreads onto the root vowel provided the first root consonant is not a voiceless obstruent (37c). Sonorants (37b) and voiced obstruents (37a) are transparent for L-tone spreading.

(37) *Nupe L-spread -- sonorants and voiced obstruents pattern together*

a.	/ɛ+dú/ →	[ɛdũ]	'taxes'
	/gí/ →	[à gĩ]	'will eat'
b.	/ɛ+lé/ →	[ɛlě]	'past'
	/lá/ →	[èlǎ]	'is carrying'
c.	/ɛ+tú/ →	[ètú]	'parasite'
	/tí/ →	[ètĩ]	'is hooting'

Such data confirm the dual status of sonorants, which depending on the [voice] specifications, can either pattern alongside the voiced obstruents, or alongside the voiceless ones. This is true for both tone (Bade vs. Nupe) and stress (Karo vs. Arabela and Pirahã).

Furthermore, one may wonder why only a certain piece of the sonority hierarchy is relevant for onset moraicity, whereas for codas it is another (bigger) chunk that roughly separates moraic sonorous segments (e.g. glides, liquids, nasals) from non-moraic non-sonorous ones (i.e. obstruents). Again this is only natural if the 'tone-and-stress pitch perturbation' phenomenon is unitary and if voicing is the only feature that can be singled out from the sonority hierarchy as responsible for such pitch effects. Other features, such as manner ones, are inert.

However, a word of caution is in order. While I would like to make this strong connection between pitch perturbation due to voicing and stress, laryngeal properties other than voicing can also be relevant for stress purposes, but in a systematically different way than the one sketched above. Thus, stronger aspiration is observed in stressed syllables in many languages including English, German, Farsi and Maori (González 2003). Glottalization (usually post-glottalization) also occurs before stressed syllables in Coast Tshimshian, Gitksan, Saanich and Lilloet (González 2003). Notably though these patterns only exhibit laryngeal manipulations found in stressed syllables. What we would like to see is cases where aspiration or glottalization attract stress in

analogy to the voicing effect found in Pirahã, Arabela and Karo. I have been unable to find such examples<sup>17</sup>, which is why voicing has a special status.

In this section, I have presented one of the main proposals in this thesis, namely that pitch perturbation caused by voicing may not only be used for tone (as standardly assumed), but also for stress, as argued for in Pirahã, Arabela and Karo (Chapter 2). Since we know that voiceless obstruents increase the pitch of the following vowel and since H-toned moras in some languages attract stress, it falls out that syllables where vowels are preceded by voiceless obstruents may be able to attract stress more than syllables with voiced obstruent onsets, which lower the pitch of the following vowel.

In addition, I have shown that sonorants present a dual status. Sometimes, they act like voiceless obstruents in attracting stress (Karo) and sometimes like voiced obstruents in being non-stress-attracting (Pirahã, Arabela). To formally capture this behaviour, I follow Rice and Avery (1989) in claiming that sonorants are inherently specified as being spontaneously voiced [SV]. The feature [voice] is something extra and not necessary for sonorants. If they have it though, then they behave like voiced obstruents which are obligatorily [voice]. If they lack it, then they can behave like voiceless obstruents, which also lack a [voice] specification. Whether the feature [voice] is present for sonorants, is language-specific. I claim that sonorants in Pirahã and Arabela are [voice], whereas sonorants in Karo are not.

#### 1.4.1.2 *Generating moraic onsets in coerced weight*

We can now move onto a more technical question, namely on how we can generate languages that have moraic onsets due to coerced weight. It has been argued that onsets lacking a [voice] specification are universally preferred to be moraic over those that possess such a specification. For this reason, the following fixed hierarchy holds.

(38) *Onset Moraic markedness*

$*\mu/\text{ONS}/[\text{voi}] \gg * \mu/\text{ONS}$

Of course this scale on its own does not tell us anything about cases where onsets are indeed moraic, simply because it bans onset moraicity altogether. For this reason, we

---

<sup>17</sup> One possible counterexample, which nonetheless occurs in a coda position (where additional properties may be also relevant) is Kwakwaka where non-glottalized sonorants act as moraic codas to the exclusion of glottalized sonorants and obstruents which are non-moraic (Zec 1988, 1995).

need another markedness constraint that imposes moraicity of this sort<sup>18</sup>. It seems that a simple extension of the WEIGHT-BY-POSITION constraint yields this effect (Hayes 1989, Kager 1999 among many others). Recall that Morén argues that coerced weight is achieved due to the force of a higher-ranked constraint that requires moraicity of segments, such as FTBIN or WBYP. Such constraints are part of the cover constraint MORAIC that Morén informally uses to assign moraicity on codas.

Although WBYP is generally well-accepted, no good justification for its existence is available (other than its practical convenience). It imposes moraicity on codas and is really an alias for MORAIC CODA which has been used in some works such as Broselow, Chen and Huffman (1997). In principle however there is no reason why a similar constraint could not apply for onsets. In fact it is merely a stipulation that it applies on codas only. For this reason, it seems that we can actually propose a constraint that assigns moraicity to onset segments too.

(39) MORAIC ONSET: Onsets are moraic

However, following Morén, I will too use the cover constraint MORAIC as the constraint which assigns moraicity on segments (unless it is required to be more precise). Now if this constraint is interleaved among the constraints in (38) we generate three possible onset weight patterns.

(40) *Patterns of coerced onset weight*

- a. \* $\mu$ /ONS/[voi] >> \* $\mu$ /ONS >> MORAIC
- b. \* $\mu$ /ONS/[voi] >> MORAIC >> \* $\mu$ /ONS
- c. MORAIC >> \* $\mu$ /ONS/[voi] >> \* $\mu$ /ONS

(40a) is really usual and familiar. It is the case where no moraic onsets of any type are allowed. The overwhelming majority of languages make use of this ranking. (40b) represents instances where only onsets that lack the [voice] feature act as moraic. Onsets

---

<sup>18</sup> The use of markedness rather than faithfulness bears on the issue that faithfulness would need to make crucial reference to the moraicity of the input. This would be undesirable for two reasons. First, we would need to restrict inputs in particular ways against Richness of the Base (cf. Prince and Smolensky 1993/2004), but could also potentially predict different outputs for the same input depending on whether underlying moraicity is assumed or not (an illustration of this problem appears when considering compensatory lengthening in §5.2). Moreover, here we are talking about coerced weight, i.e. weight that appears in the output, therefore recourse to faithfulness would not be able to help us achieve the right results. For a similar argument, see Morén (1999/2001).

marked as [voice] fail to do so. Depending on whether sonorants are specified as [voice] or not, through this pattern we not only generate Pirahã and Arabela (where sonorants are [voice], thus only voiceless obstruents are moraic), but also Karo (where sonorants lack [voice], therefore both voiceless obstruents and sonorants are moraic).

The last pattern (40c) refers to languages where all onsets are moraic independently of their type. Bella Coola Word Minimality offers a potential case for such pattern (cf. (14)). As we will see in Chapter 4, my claim is that in this language, moraic onsets arise only in [CV] words, so that the bimoraic word minimum is satisfied. This can be equally satisfied though by all types of onsets, e.g. [ʔi] ‘fast’, [xʷ] ‘to pull’, [w] ‘to spill’, [mu] ‘fishtrap’ (Nater 1979: 170, 171, 175), suggesting that sonorant onsets too are moraic. Recall however that I have been arguing that sonorants may either be specified as [voice] or lack such a feature altogether, thus to be able to argue that Bella Coola offers evidence for (40c), we would need to see what happens with voiced obstruents which are consistently [voice]. Had Bella Coola allowed voiced obstruents occurring in [CV] words, then we would be bound to argue that they are moraic, in the same way all other onsets are, supporting (40c); on the other hand, if the language had voiced obstruents, but these failed to occur in [CV] words, then presumably voiced obstruents would be unable to be moraic onsets, thus only segments lacking [voice] could be moraic (evidence for (40b)). Unfortunately, since Bella Coola lacks word-initial voiced obstruents, both options seem possible<sup>19</sup>.

#### 1.4.1.3 *Moraic onset geminates*

I have argued that there are two types of moraic onsets: those which are the result of coerced weight as in Pirahã or Arabela and those which are the result of distinctive weight as in Pattani Malay or Trukese. In this section, I explore the latter case.

Distinctive weight distinguishes between singleton and geminate consonants. Traditionally (after Hayes 1989), geminates have been assumed to be consonants which are underlyingly specified with a mora. Thus, the distinction between singletons and geminates is one between input non-moraic consonant and input moraic consonant.

---

<sup>19</sup> Alternatively, one could seek evidence whether sonorants are [voice] or not through other processes in the language, but I have been unable to do so.

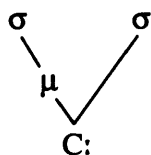
(41) *Singletons vs. geminates in the input (Hayes 1989)*



Since then, various works have questioned the validity of the claim that geminates are always moraic including most prominently Tranel (1991) and more recently Muller (2001). Others however show that instances supporting the absence of moraicity in geminates may be refuted or at the very least be weakened, once data are examined more closely, concluding that geminates are consistently moraic (Curtis 2003). I will concur with the latter approach and argue that there also exist geminates which occur in the onset position and whose primary characteristic is to be moraic, i.e. they are moraic onsets. I will also address cases like *Tukang Besi* or *Berawan* which have been argued to possess onset geminates (Chapter 7), and show that these seem to require syllabification in the onset position, but provide no evidence for onset moraicity, thus they are better represented as ‘doubled consonants’ (see below, Ham 2001 and §6.5). My attention will thus be focused on languages where geminate moraicity is supported and where geminates need to be syllabified wholly in the onset.

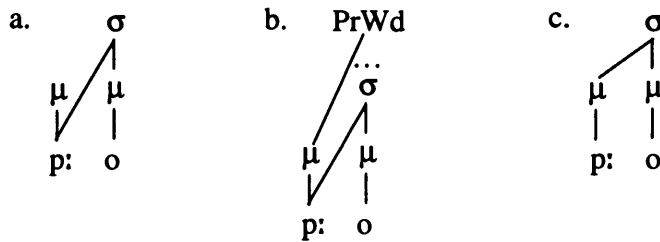
With this disclaimer in mind, it is worth considering why and how onset geminates can be possible. First of all, recall that the standard representation of *word-medial* geminates is the familiar flopped structure where the first part of the geminate syllabifies in a coda (and carries a mora), and the second part directly associates to the onset of the following syllable. This structure achieves two goals: i) it represents the increased length of geminates compared to singletons and ii) it allows the coda, rather than the onset, to host a mora.

(42) *Flopped structure of geminates word-medially*



A major problem that such a configuration faces is the representation of initial geminates. In particular, no coda is available that can simultaneously host the mora and the first part of the geminate, effectively making the representation of an initial geminate practically impossible without any further modifications. Curtis (2003) addresses this issue and presents a number of possible solutions.

(43) *Possible moraic representations of initial geminates*



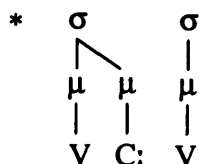
Curtis uses Trukese as her example of such representations. In Trukese, there is a bimoraic word minimum. The fact that C:V words are accepted indicates that geminates indeed contribute a mora. (43a) is the structure of initial geminates proposed in Davis (1999), which is correctly abandoned. The reason is that WdMin can only be satisfied if the mora is linked to a higher structure in order that it is counted. Here it is unaffiliated, thus this representation loses the insight that the mora counts for minimality purposes. (43b) is the structure Curtis chooses, but this poses different problems. Apart from the double association this geminate bears, it otherwise looks very much like an unsyllabified consonant like the ones surfacing in Bella Coola (Bagemihl 1991), Piro (Lin 1997) or Arabic (Kiparsky 2002). Although these are not dominated by a syllable, they can still be licensed by a mora linked to the foot or to the PrWd. This is what happens in (43b) too, but obviously the status of geminates seems to be different. No evidence for absence of syllabification is available, therefore this representation has to be discarded too. Curtis finally entertains the possibility of the structure in (43c) - i.e. an onset geminate - but does not investigate it further.

This, I claim, is actually the right representation of initial moraic geminates. Moreover, I propose that this representation is not only possible word-initially, but also word-medially as languages like Marshallese seem to suggest (§6.4.2).

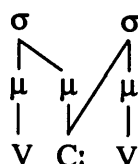
But how about the flopped structure characteristic of geminates? This can no longer be maintained. To answer this problem, I follow Ham (2001), who insightfully observes that: "...moraic theory itself does not necessitate the 'flopped' representation of medial geminates; this is required instead by the dispreference for onsetless syllables" (2001: 13). In other words, double linking of the geminate is not required by moraic theory. All is required is that the geminate bears a mora. The fact that it also appears in the onset position is the result of preferring more well-formed syllables, i.e. those beginning with an onset (compare onsetless (44a) with onsetful (44b)).

(44) *Geminates word-medially*

a. Avoidance of onsetless  $\sigma$ s

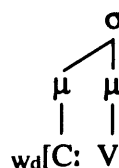


b. Preferred structure

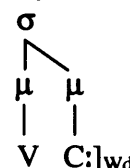


This of course has certain repercussions for geminates in other positions. Single linking and absence of flopped structure is the default configuration for initial and final geminates as shown below.

(45) a. Word-initially (proposed here)



b. Word-finally (Ham 2001)

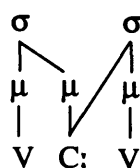


But of course this representation of final geminates is now no longer different from the one imposed for singleton moraic consonants. To distinguish between the two, Ham makes the strong prediction that there is no language that has both final geminates and CVC singletons which count as heavy<sup>20</sup>. Ham claims that this is universally true, but extensive testing is required. His examples include Bernese, Levantine Arabic and Hungarian, which treat final CVC singletons as extrametrical. One can thus claim that MORAIC with respect to codas does not apply word-finally. Despite that, final geminates are heavy. This can be explained because their weight is underlying, so even if MORAIC fails to apply word-finally, these consonants are already moraic.

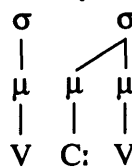
Given the proposals of the current work, an extension to Ham's idea is natural. More specifically, since moraic onsets are explicitly acknowledged, it is possible that the structure assumed in (45a) can also occur word-medially. The two possibilities, the flopped structure and the moraic onset geminate, are shown in (46).

(46) *[VC:V] sequence word-medially*

a. Avoidance of moraic onsets



b. Avoidance of syllables with codas



<sup>20</sup> However in this language, non-final singleton CVCs can be heavy.

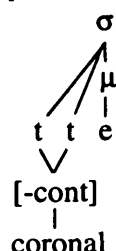


Languages may choose between (46a) and (46b). Importantly, the former renders the first syllable heavy, while the latter has the same effect for the second syllable. (46a) does better in having no moraic onsets at all, which entails that the constraints banning all types of moraic onsets - let us call this generically \*MORAIC ONSET - are highly-ranked. (46b) is better in avoiding extra codas. Assuming a language has \*MORAIC ONSET low-ranked and provided NOCODA >> \*MORAIC ONSET, the latter constraint will prefer a structure where codas are minimised by virtue of tautosyllabic assignment in onsets (as in (46b)). This ranking of NOCODA - if sufficiently low-ranked - does not entail that codas will be banned altogether, but that they will be avoided wherever possible. However, since most languages rank \*MORAIC ONSET highly, it makes sense that the majority of languages syllabify geminates heterosyllabically (46a) rather than tautosyllabically (46b). The latter is merely rarer, but by no means impossible.

Empirical evidence for (46b) appears in Marshallese (§6.4.2), where default stress is leftmost in a trisyllabic window at the right edge, e.g. *ékajet* ‘to judge’ (this also shows that singleton codas are not moraic), unless the word includes a geminate in which case stress docks on the syllable that *follows* it, e.g. *jibbúŋ* ‘morning’. I will argue that this is only possible if the geminate is moraic and wholly syllabified in the onset.

Ham also considers geminates which seem to lack moras and claims that while all geminates are moraic, not all long consonants are geminates. In his opinion, what are dubbed ‘weightless’ geminates are simply doubled consonants with two root nodes. To this point, he adopts Hayes’ representation of fake geminates which arise under morphological concatenation. These involve a complex onset made of two identical consonantal root nodes (see also Selkirk 1990).

(47) *Representation of fake geminates and (possibly) of long non-moraic onset Cs*



One then expects that geminates and doubled consonants may co-occur in the same language, which is what Ham argues for the case of Bernese. In all likelihood, this may be the correct representation for alleged geminates in the languages to be examined later on, such as Berawan (García-Bellido and Clayre 1997; §7.3.1). This is particularly

plausible if one does not necessarily correlate the presence of fake geminates with morphological pressures, but instead sees them as the accidental phonological combination of alike consonants.

## **1.5 Organisation of the thesis**

This introductory chapter has set the scene of what will follow. Contrary to most theories of weight, I have argued that onset weight by means of moraic onsets is possible. Acknowledging this fact allows us to explain numerous data and phenomena which up to now remained obscure. Onset weight however is not unrestricted. It either takes the form of coerced onset weight or distinctive onset weight (Morén 1999/2001).

I start by considering the first, i.e. where onsets appear as moraic only on the surface. In Chapter 2, I use data from stress in Pirahã, Arabela and Karo to show that in coerced weight, only a particular subset of segments can carry moraicity, depending on the presence or absence of the feature [voice], whose pivotal role in relationship to pitch, tone and stress I have discussed in §1.4.1.1. Chapter 2 thus mainly deals with languages where the ‘quality of the onset’ influences stress assignment. There is another aspect to onset-sensitive stress, namely the effects relating to the ‘presence of the onset’. These are discussed in Chapter 3 by making reference to a handful of Australian and Native American languages, as well as Dutch. Crucially, I argue that moraic onsets are only relevant in the ‘quality of the onset’, but not in the ‘presence of the onset’ effects.

Despite the numerous deficits that prominence approaches are confronted with (§1.3), I do acknowledge that these can nevertheless account for the stress data presented in Chapters 2 and 3, since prominence is particularly well-suited to deal with this phenomenon. What it finds impossible to explain however, are other phenomena, where it is beyond doubt that utilising weight is in order. Thus, finding data where onsets are computed in weight calculations provide a very strong argument for the existence of moraic onsets. Indeed such cases exist as I argue in Chapters 4-6.

In Chapter 4, I investigate Word Minimality in Bella Coola and show that [CV] words behave like other minimal words in the language which happen to be bimoraic. The natural conclusion then is that these are bimoraic too. Chapter 5 looks at compensatory lengthening in Samothraki Greek, where the deletion of an onset results in lengthening of the following vowel.

Chapter 6 has a dual purpose; on the one hand it continues and extends the argument in favour of moraic onsets, but also introduces the notion of distinctive onset weight, i.e. lexically-specified weight, in the guise of (initial) geminates. Here, I argue

that: i) if the right representation of a geminate is to be underlyingly moraic, and ii) if moraic theory does not necessitate the flopped structure of geminates, then given the existence of onset weight, we are bound to find geminates which are best represented as moraic onsets. I claim that this is exactly right and straightforwardly accounts for the long-standing problem of initial geminates, as in Pattani Malay and Trukese. Importantly, this representation can also be extended word-medially, as in Marshallese.

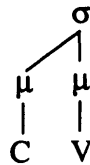
Chapter 7 again has two aims. First, it lists a fair number of languages which could be taken to support onset moraicity through a number of processes and phenomena, e.g. reduplication, gemination, metrics, etc., but the data are not yet robust enough to do so conclusively. Its second aim is equally important; it investigates languages which in previous accounts have implicitly or explicitly been argued to support onset moraicity, but where I argue that this is false. This chapter then can serve as a starting point on what kind of patterns we can expect and what we should be aware of in our search for onset weight languages.

Finally, in Chapter 8, I provide a summary of the most important points of the thesis and then venture out on a more exploratory track. In particular, I report on the results of a tentative experiment I conducted to test the hypothesis that the pitch perturbation caused by (the lack of) voicing can affect stress. I then point out directions for future research including the relationship between CCV and CV syllables when onset weight is relevant, the possibility of tone linked to moraic onsets and the necessity for further language documentation or their re-examination in the light of the proposed modified syllable and weight model that allows both:

(48) a. *Non-moraic onsets*



b. *Moraic onsets*



## Chapter 2

### Onset Moraicity and Stress

#### 2.1 Introduction

I begin the exploration of weightful onsets by examining onset-sensitive stress. Recall from the previous chapter that we have claimed that onset-sensitive stress depends on two independent dimensions: a) the presence of the onset, b) the quality of the onset.

(1) *Presence and quality of onset interaction in stress*

		Quality of onset	
		YES	NO
Presence of an onset	YES	Pirahã (Arabela) <sup>1</sup> (❶)	Aranda, Banawá, Dutch (❷)
	NO	Karo (Arabela) (❸)	Standard Greek, Russian, etc. (❹)

Effects on stress due to onset presence (patterns ❷ and ❶) are found in some Australian (Aranda, Alyawarra) and American Indian languages (e.g. Banawá, Iowa-Oto), where the first onsetful syllable receives stress. Unlike these languages, where such effect only appears at the edge, in Pirahã the presence of an onset attracts stress in other positions too, so that (C)VV > VV. However, I agree with Goedemans (1998) in claiming that this phenomenon is a matter of alignment of a stressed syllable with an onsetful one, and not one of onset weight. While I will return to this issue in detail in Chapter 3, I will now focus on what I consider as genuine effects of onset sensitivity with respect to weight (patterns ❸ and ❶).

These I claim are evident in at least three languages: Pirahã, Arabela and Karo. In those languages, onset consonants that lack a voice feature attract stress more than those which have a [voice] feature. I start the discussion with data from Pirahã.

---

<sup>1</sup> Arabela most likely lacks onsetless syllables, therefore it cannot serve as a testing ground with respect to the presence of the onset issue. This is why I position it in both cells.

- (2) *Pirahã examples* (in Pirahã I mark stressed syllables in bold; acute accent = H tone, no accent = L tone; E&E=Everett&Everett 1984, E=Everett 1988)

ko.ʔo.pa	“stomach”	[E: 239]
ti.po.gi	“species of bird”	[E&E: 710]
ʔa.ba.gi	“toucan”	[E&E: 710]

While default stress in Pirahã is rightmost, e.g. *koʔopa*, it can move away from that position if the word consists of equally heavy syllables which only differ with respect to the voicing of their onsets. A syllable with a voiceless onset attracts stress, which is why we get *ʔabagi* and not *\*ʔabagi*.

In Arabela (Payne and Rich 1988), stress is generally rhythmic (3a), with the exception of cases like those in (3b). There, rhythmic stress, as depicted in the second column is unacceptable. Instead, primary stress moves from the ultima to the penult, but only when the penult contains a voiceless consonant and the ultima a voiced one. This evinces the priority of voiceless onsets over voiced ones on stress.

- (3) *Arabela stress*

a.	tènakári		‘afternoon’
	sàmarú		‘spirit’
b.	nòwaʃíjano	*nòwaʃíjano	‘brightened’
	sàpohòsano	*sàpohòsano	‘deceived’

There is a reason why I choose to make reference to the contrast above (cf. §1.4.1.1) as one between [voice] consonants and those that lack this feature. This is because sonorants vary from language to language as to whether they pattern alongside voiced or voiceless obstruents. In the languages where sonorants pattern with voiced obstruents, we assume that sonorants are [voice] too, but in the cases where they do not, they lack this feature altogether. Of course sonorants always have the [SV] (spontaneous voice; cf. §1.4.1.1) feature that distinguishes them from obstruents as claimed by Rice and Avery (1989). In Arabela, it is clear that sonorants behave like voiced consonants<sup>2</sup>. In Pirahã, no published data confirm this either way. However, Everett (p.c. 18/03/04, 03/06/04 and 20/10/05) provides a clear answer to this question. He notes that there are no input sonorants in Pirahã, but they may arise as allophones of the voiced stops. The phonemic inventory of Pirahã consists of /p/, /t/, /ʔ/, /s/, /h/, /b/,

<sup>2</sup> Rich (1963) reports that Arabela also has voiced stops and voiced fricatives as allophones of the voiceless stops. Unfortunately no data of the type in (3b) involve voiced obstruents, but I would anticipate that they behave like the sonorants.

/g/, /i/, /a/ and /o/. Input /b/ has three surface realizations: [b], [m] and the bilabial trill [B], where the last two are sonorants. Input /g/ emerges as [g], sonorant [n] or as the sui generis voiced apico-laminal alveolar-labial double flap [r̥] only reported in Pirahã. Pirahã is nowadays the sole survivor of the Mura language family. While [n] seems an unlikely allophone for /g/, this seems more reasonable in the light of the fact that Proto-Mura \*d shifted to /r/ in Mura and /g/ in Pirahã (Everett p.c. 18/03/04).

Some rules that regulate the distribution of the allophones of the voiced stops are given below:

- (4) b → m/ pause \_\_
- (5) b varies freely with B/ i or a\_\_o
- (6) g varies freely with [r̥]/ o\_\_i

According to E&E (1984: 708, 710), the following words are stressed as shown:

- (7) a.      **ʔibogi**           ‘milk’
- b.      **biísai**           ‘red’

Each of these fulfils the environment where at least one of the rules above may apply.

- (8) a.      (5)/(6) apply           **ʔiBogi / ʔiBoŕ̥i / ʔiBoŕ̥i**
- b.      (4) applies           **miísai**

The crucial observation is that in each of these cases, where a sonorant is involved, stress is identical to the one where voiced stops appear instead. Therefore, we can safely assume that sonorants are specified as [voice] in Pirahã.

Still, in other cases, sonorants do not behave like voiced obstruents. In particular, Karo stress is word-final when the final onset is a voiceless obstruent or sonorant (9a), but when the final syllable contains a voiced obstruent onset, then stress is attracted to the penult provided it contains an onset which is either voiceless (9b) or sonorant (9c) (if the penult has a voiced onset too, stress stays put on the ultima).

- (9) *Karo onsets and stress* (Gabas 1999: 39-41; stress in bold)

- a.      **nahe**k****           ‘fontanel’
- kiriwep**           ‘butterfly’
- b.      **cigi**           ‘spot’
- pibe?**           ‘foot’
- c.      **yɔgɔ**           ‘eel’
- məga**           ‘mouse’

We would thus say that in Pirahã and Arabela, sonorants are specified as [voice], whereas in Karo they are not. The fact that [voice] is not a necessary feature for sonorants allows them to manifest the dual nature they present. Similar effects originating in voicing have also been traced in tone, which led us to the proposal in the previous chapter that pitch perturbation due to voicing can be either construed as tone or as stress. For the stress facts under consideration, in §1.4.1.1 I have argued that the appropriate explanation lies in making use of moras for onsets. The current chapter offers detailed analyses of the three languages discussed above.

## 2.2 Pirahã

### 2.2.1 Data and generalisations

Pirahã is spoken by about 200 Indians in northwestern Brazil. Before moving to the extremely complicated weight scale for which Pirahã is notorious, it is worthwhile mentioning some background on the language. The Pirahã syllable structure is (C)V(V), which means that codas are not permitted. Single monovocalic syllables, i.e. V, are not admitted either. As mentioned previously, the language also has tone, but this does not interfere with stress (K. Everett 1998). All vowels are specified for tone (´ = H tone, no accent = low tone). For convenience, the relevant data which show the independency of tone and stress are repeated here.

(10) *Pirahã: tone independent of stress* [P=voiceless onset, B=voiced onset]

- |     |                 |                |
|-----|-----------------|----------------|
| i)  | <b>PVBV</b>     |                |
| a.  | <b>HH:</b> tígí | ‘small parrot’ |
|     | <b>LL:</b> pígi | ‘swift’        |
| b.  | <b>LH:</b> sabí | ‘mean’         |
|     | <b>HL:</b> ?ábi | ‘to stay’      |
| ii) | <b>PVPV</b>     |                |
| a.  | <b>LH:</b> ti?í | ‘honey bee’    |
|     | <b>HL:</b> tíhi | ‘tobacco’      |

We can now examine the language’s weight hierarchy in detail and its relation to stress (Everett&Everett 1984, Everett 1988).

(11) *Pirahã weight scale*

PVV > BVV > VV > PV > BV

The following data illustrate this scale. The stress algorithm is basically ‘stress the rightmost heaviest syllable within the trisyllabic window at the right edge of the word’. Pirahã, like other languages in the world, e.g. Greek, Italian, Spanish, only treats one of the last three syllables as possible stress bearers. (12e) in particular illustrates that when syllables are of the same weight, then the rightmost one receives stress. For example, in *tipogi*, both the antepenult and the penult can carry stress by virtue of being the heaviest in the word. Yet, only *po* does so, since it is the rightmost of the two.

(12) *Pirahã examples* (E&E=Everett&Everett 1984, E=Everett 1988)

- |    |                              |  |   |
|----|------------------------------|--|---|
| a. | PVV > BVV                    |  |   |
|    | <b>káo</b> .bá.baí           | “almost fell”                            | [E: 239]                                      |
|    | pa. <b>hai</b> .bí           | “proper name”                            | [E&E: 708]                                    |
|    | píi.bi.gái                   | “deep water”                             | [Everett p.c.]                                |
| b. | BVV > VV                     |  |   |
|    | bíi.oá.ii                    | “tired [literally: being without blood]” | [Everett p.c.]                                |
|    | poo. <b>gái</b> .hi.aí       | “banana”                                 | [E&E: 709]                                    |
| c. | VV > PV                      |  |   |
|    | pia.hao.gi.so. <b>ai</b> .pi | “cooking banana”                         | [E&E: 710]                                    |
| d. | PV > BV                      |  |   |
|    | ? <b>a</b> .ba.gi            | “toucan”                                 | [E&E: 710]                                    |
|    | ti. <b>po</b> .gi            | “species of bird”                        | [E&E: 710]                                    |
| e. | rightmost heaviest stress    |  |   |
|    | ? <b>a</b> .ba. <b>pa</b>    | “Amapá” (city name)                      | *? <b>a</b> .ba.pa [E&E: 710]                 |
|    | ho.áo. <b>fi</b>             | “shotgun”                                | *ho.áo.fi [E&E: 710]                          |
|    | ti. <b>po</b> .gi            | “species of bird”                        | * <b>ti</b> .po.gi [E&E: 710]                 |
|    | paó.hoa. <b>hai</b>          | “anaconda”                               | *paó.hoa.hai/* <b>paó</b> .hoa.hai [E&E: 707] |

Observation of these data leads to the conclusion that weight depends on three different parameters, laid out below.

(13) *Pirahã weight depends on*

- |             |              |                |
|-------------|--------------|----------------|
| i) VV vs. V | ii) CV vs. V | iii) PV vs. BV |
|-------------|--------------|----------------|

The first parameter recognises the heaviness of long vowels over short ones. This is a very familiar and recurrent pattern cross-linguistically. The second factor refers to the ability of an onsetful syllable to attract stress more than an onsetless one. Such an effect is not too uncommon (cf. numerous Australian and a few American Indian languages),



and can be attributed to the preference of aligning stressed syllables with onsetful ones (Chapter 3). Thus my focus for the ensuing analysis is going to be on the third factor, namely sensitivity to onset quality.

### 2.2.2 Analysis

The property that renders Pirahã unusual is the fact that voiceless obstruent onsets attract stress more than those with sonorant or voiced stop ones. The functional explanation of this distribution relates voicing with stress in analogy to the effects that voicing has on pitch in tone languages. Formally, the distinction is attributed to the universally fixed hierarchy:

- (14) *Onset moraic markedness*  
 $*\mu/\text{ONS}/[\text{voice}] \gg * \mu/\text{ONS}$

The ranking above expresses the fact that [voice] moraic onsets are less preferable than those which lack such a feature. Of course such a ranking makes some sense if there is another constraint that enforces moraicity on onsets. As we have seen, in line with Morén (1999/2001), such a constraint can be named MORAIC and includes constraints that require moraicity on certain occasions, such as WDMIN, WBYP and others. In the same way that codas may acquire moraicity in the output due to WBYP, I claim that a similar effect arises with onsets. A more uniform representation of these constraints would make use of MORAIC ONSET to assign moraicity to onsets and MORAIC CODA bringing the same effect on codas. To account then for the distribution of moraic onsets in Pirahã, we only need the ranking below [I use the more generic constraint MORAIC for convenience].

- (15) *Pirahã moraic onsets*  
 $*\mu/\text{ONS}/[\text{voice}] \gg \text{MORAIC} \gg * \mu/\text{ONS}$

To illustrate, let us see how moraicity would be assigned to a word like /tipogi/. Note that I assign moras onto input vowels following Hayes (1989) who considers short vowels as underlyingly monomoraic, whereas long vowels as underlyingly bimoraic<sup>3</sup>.

---

<sup>3</sup> This is a common assumption in most analyses. Even Rosenthal (1994) who argues that vowels acquire moras through the constraint V-mora, claims that long vowels are lexically specified in terms of two

While the inclusion of moras on vowels here makes the tableaux look more complex, it is a necessary complication given that the language not only distinguishes between short and long vowels, but this distinction also proves important in the understanding of weight and stress assignment. Observe that superscripted moras after an onset indicate that it is a moraic onset.

(16) \* $\mu$ /ONS/[voice] >> MORAIC >> \* $\mu$ /ONS for /tipogi/

	$ti^{\mu}po^{\mu}gi^{\mu}$	* $\mu$ /ONS/[voice]	MORAIC	* $\mu$ /ONS
a.	$t^{\mu}i^{\mu}.p^{\mu}o^{\mu}.gi^{\mu}$		*	**
b.	$t^{\mu}i^{\mu}.po^{\mu}.gi^{\mu}$		**!	*
c.	$t^{\mu}i^{\mu}.p^{\mu}o^{\mu}.g^{\mu}i^{\mu}$	*!		**

The last candidate leaves the competition early on, since by having moras on all onsets, it violates \* $\mu$ /ONS/[voice] due to  $g^{\mu}i^{\mu}$ . The second candidate avoids this problem by having the onset of  $gi$  moraic, but also fails to assign a mora to one of the voiceless onsets. As a result, it violates MORAIC more severely than (a) which manages to assign moras on all voiceless obstruent onsets.

The proposal that voiceless onsets are heavier than voiced ones also receives some external empirical support. First, there is a tendency of voiceless obstruents to be longer than voiced ones (Ohala 1983, 1997, Maddieson 1997). This property is phonologized in Swiss German (Ham 2001 for Bernese in particular, Kraehenmann 2001) and appears as an underlying phonemic contrast. More generally, it is the case that in languages that allow only one type of geminates, these are commonly voiceless rather than voiced, e.g. Lak, Nez Perce, Ocaina, Ojibwa, Totonac, Yakut and Japanese (Morén 1999/2001).

Consonant gradation in Finnish and related languages can be understood along those lines too (Anttila 1994, Gordon 1998). In Finnish, double stops degeminate and simple stops weaken before a branching rime, so that we get alternations of the type  $tt > t$  and  $t > d$ , e.g. *takka > takan* ‘fireplace-GEN’, but *takkana / \*takana* ‘fireplace-ESS’ as there is no branching rime there, and similarly *suti > sudin* ‘brush-GEN’, but *sutina / \*sudina* ‘brush-ESS’. Also in Yolngu Djapu, Morphy (1983) mentions that the contrast between voiceless and voiced stops has also been described as one between geminate and simple stops respectively, so /b/ actually corresponds to  $p$  and /p/ to  $pp$ . In the Kuna “talking backwards” language game (Sherzer 1970), the first syllable of the word moves

---

moras and are not subject to V-mora. I take a more simplified approach in assuming that both short and long vowels are specified as such in the input by means of moras.

to the end. The inversion is straightforward when the consonants involved include voiced stops e.g. *dage* → *geda*, *obsa* → *saob*, but when voiceless stops are encountered, then the inversion takes a different shape. Thus for the regular form *sapan* we do not get *\*pansa* as anticipated, but instead *bansab* and similarly *sate* → *desad* and not *\*tesa*. This is very similar to Djapu above. If voiced stops actually act as singletons, whereas the voiceless ones as geminates, then this is exactly the pattern we would expect to find.

All these facts make it plausible that voiceless obstruents can be phonologically heavier as onsets than their voiced counterparts, because they are phonetically longer too<sup>4</sup>. Onset moraicity of course has repercussions on the moraic make-up of Pirahã syllables, which is illustrated in (17).

(17) *How many moras per syllable?*

	Total number of moras	Onset	Mora composition
PVV =	3μ	YES	(2μ-nucleic, 1μ-onset)
BVV =	2μ	YES	(2μ-nucleic)
VV =	2μ	NO	(2μ-nucleic)
PV =	2μ	YES	(1μ-nucleic, 1μ-onset)
BV =	1μ	YES	(1μ-nucleic)

Two things seem clear at a first glance. PVV is unambiguously heavier than all other syllables, in the same way that BV is the lightest. The former is trimoraic, whereas the latter is only monomoraic. A potential problem now is that the remaining syllable types, BVV, VV and PV are all considered bimoraic. Thus they should behave as equally heavy, and yet a well-defined hierarchy exists among them, i.e. BVV > VV > PV. To see why this should be the case, we need to take into account that BVV differs from VV in having an onset that the latter lacks. Such preference of stressing onsetful syllables can be captured by an alignment constraint that matches stress with onsetful syllables.

(18) **ALIGN<sub>FO</sub>**: Align-L (σ, C) [cf. Goedemans' (1998) AlignFtO and Chapter 3]

This leaves us with the VV > PV part of the scale. PV not only has an onset, which is missing from VV, but it also happens to be a voiceless one, hence moraic. Nonetheless, it acts as lighter than VV. Observe however that VV has two moras too, but both come from the nucleus, unlike PV where one is linked to the nucleus and the other to the onset. In other words, this pattern indicates that nucleic weight takes priority over the

<sup>4</sup> As far as Pirahã goes, experimental work by K. Everett (1998) shows that onset consonants in stressed syllables are longer than those in stressless syllables, but no discussion is made about the voicing properties involved, and whether these have any effect.

overall weight. Translating this in terms of constraints, we need the constraint which requires that heavy syllables receive stress (i.e. the Weight-to-Stress Principle / WSP cf. (19)) to appear in a special version where heavy syllables due to nucleic moras receive stress. Such a constraint can be called WSP(N) where N stands for nucleus (20).

(19) **WSP:** Heavy syllables are stressed (N.B: gradiently assessed, i.e. a stressless *bimoraic* syllable incurs *one* violation, a stressless *trimoraic* syllable incurs *two* violations)

(20) **WSP(NUCLEUS):** Heavy syllables due to nucleic moras are stressed

Similar constraints have been proposed in Hammond (1999) with respect to English, where it is claimed that under certain circumstances CVV behaves as heavier than CVC. In fact, in the common case where codas count for weight purposes, although many languages treat CVV and CVC as equally heavy, e.g. Brahui, Latin, Votic (Gordon 1999), there are numerous others which observe a CVV > CVC weight hierarchy, e.g. Kashmiri (Morén 1999/2001, 2000), Klamath and Chickasaw (Gordon 2002).

It is the combination of onset moraicity (15), the alignment of a stressed syllable with an onsetful one (18) and the WSP facts (19-20) that produce the complex system of Pirahã. One final ingredient is missing; the fact that among equally heavy/light syllables (12e), the rightmost one receives stress. Another alignment constraint accounts for that<sup>5</sup>.

(21) **ALIGN-HEAD-RIGHT:** Align the head syllable of a prosodic word to the right edge of the prosodic word [McCarthy and Prince (1993)]

Putting all this together, the following ranking obtains.

(22) *Proposed ranking for Pirahã:* WSP(N), WSP >> ALIGN<sub>HO</sub> >> Align-Hd-R

Several ranking arguments lead us to this conclusion. These are shown presently.

---

<sup>5</sup> We also need an extra constraint that gives us the trisyllabic window at the right edge. Although this issue is interesting on its own merit (cf. Green 1995 on Pirahã), for current purposes, I will set this aside and assume the \*Extended Lapse Right constraint which bans sequences of more than two consecutive stressless syllables at the right edge of the word [Elenbaas and Kager (1999), Gordon (2005)]. To simplify, in the tableaux that follow, \*Extended Lapse Right is assumed but not included.

(23) WSP >> ALIGN-HD-R

PVV.BVV: toi.bai “woman’s name” [Everett p.c.]

	to <sup>μ</sup> i <sup>μ</sup> ba <sup>μ</sup> i <sup>μ</sup>	WSP	ALIGN-HD-R
a.	t <sup>μ</sup> o <sup>μ</sup> i <sup>μ</sup> .ba <sup>μ</sup> i <sup>μ</sup>	*	*
b.	t <sup>μ</sup> o <sup>μ</sup> i <sup>μ</sup> .ba <sup>μ</sup> i <sup>μ</sup>	**!	

In a nutshell, misalignment of stress is compelled due to better satisfaction of the WSP by the first candidate compared to the perfectly aligned second one. This tableau however merits more extensive discussion. The careful reader may have noticed that violations of WSP are gradiently assessed so that (23a) incurs one violation and (23b) two. However, this is not what one would normally expect if WSP were computed in the standard way, i.e. heavy syllables must be stressed. In this interpretation, WSP only cares that heavy syllables receive stress, thus each of (23a) and (23b) should have just one violation of WSP leading to a tie. ALIGN-HD-R would thus determine the outcome, by wrongly selecting (23b) as the winner, since it presents perfect alignment. But this is not the only possible way to understand WSP. Note that the syllable [.t<sup>μ</sup>o<sup>μ</sup>i<sup>μ</sup>.] has three moras, whereas [.ba<sup>μ</sup>i<sup>μ</sup>.] only two. Given that the former is heavier than the latter, it intuitively makes sense that when comparing heavy syllables, WSP is more severely violated, if the heavier one [t<sup>μ</sup>o<sup>μ</sup>i<sup>μ</sup>.] remains unstressed.

There are at least two ways to tackle this issue. The first is to make use of PKPROM instead. This constraint, proposed in Prince and Smolensky (1993: 39, 41), indicates that:

(24) PKPROM: μμμ > μμ > μ

PKPROM directly looks into the moraic composition of syllables and picks out the most prominent for stress. If we had used it in (23) instead of WSP, then (23a) would be correctly chosen as the winner, since its stressed syllable is trimoraic and therefore has priority for stress over the second syllable. So the question is now whether we should prefer PKPROM instead of WSP. One thing to consider is that even if PKPROM replaces WSP, it cannot subsume the effects of WSP(N) which we need to include in our analysis as I will show in (25) below. Since WSP(N) is used, we expect WSP to be present in the ranking too, therefore it would be preferable to try to solve the problem by making use of this constraint.

This is possible if we assume that WSP is gradiently evaluated depending on the moraic make-up of syllables (but see McCarthy 2003b who argues against gradience). In some sense then WSP subsumes PKPROM since it can now favour stressing a super-heavy syllable over a heavy one. This is why (23a) only receives one violation of WSP, whereas (23b) receives two; the former leaves a heavy syllable unstressed, while the second leaves a super-heavy unstressed, making this a fatal violation. It is well-known that while standard WSP and PKPROM overlap to a certain extent, originally only PKPROM could select a super-heavy stressed syllable over a heavy stressed one. This is no longer the case if WSP is amended as suggested.

Nonetheless, calculation of violations is still different. WSP receives violations for every heavy/super-heavy syllable that remains unstressed, whereas PKPROM is satisfied as soon as the heaviest syllable in the word is stressed. This may be just one syllable leaving others - which are equally heavy - unstressed. For PKPROM this makes no difference. All it matters is to stress one - the heaviest - syllable in the word. This difference may prove important before deciding to discard PKPROM after the present modification of WSP, but I will leave this issue open for further investigation. With this in mind, we can move on and see why PKPROM could not replace WSP(N) too.

(25) WSP(N) >> ALIGN $\phi$ O<sup>6</sup>

PV.VV.PV: ho.af.pi “type of fish” [E&E: 710]<sup>7</sup>

	ho <sup>μ</sup> a <sup>μ</sup> i <sup>μ</sup> pi <sup>μ</sup>	WSP(N)	ALIGN $\phi$ O
a.	h <sup>μ</sup> o <sup>μ</sup> .a <sup>μ</sup> i <sup>μ</sup> .p <sup>μ</sup> i <sup>μ</sup>		*
b.	h <sup>μ</sup> o <sup>μ</sup> .a <sup>μ</sup> i <sup>μ</sup> .p <sup>μ</sup> i <sup>μ</sup>	*!	

Candidate (25a) with a stressed VV syllable is preferred over a PV stressed one. Note that neither WSP or PKPROM alone would be able to select the right winner here. Every syllable in this word is bimoraic, therefore no matter which is stressed, the results are the same. If PKPROM is used, then no candidate violates it, because the heaviest syllable in the word happens to be a bimoraic one. On the contrary, if WSP is used, then every candidate would violate it twice, since in each candidate two heavy bimoraic syllables

<sup>6</sup> Words in Pirahã must begin with consonants (Everett p.c.), so no disyllabic word of this type is available.

<sup>7</sup> One could wonder what if in this tableau, the relevant constraints were WSP(N) and ALIGN-HD-R instead. While the correct result would be generated too, we know that ALIGN $\phi$ O is independently needed in rankings like the one in (27), where it is in conflict with ALIGN-HD-R. Therefore, its relative ranking with respect to other constraints needs to be established. Had (25) been WSP(N), ALIGN $\phi$ O >> ALIGN-HD-R, then (25b) would have wrongly won due to perfect alignment. WSP(N) >> ALIGN $\phi$ O ensures that this is not the case.

stay unstressed. In both cases then, candidates would tie, and ALIGN $\sigma$ O would wrongly favour candidate (25b), since it stresses an onsetful syllable (and manages to align the stressed syllable with the right word edge). The use of WSP(N) however resolves the problem. WSP(N) focuses on nucleic moras only, which means that while *ho* and *pi* are heavy, they are not heavy due to the bimoraicity of their nucleus. Leaving them unstressed does not violate WSP(N). On the other hand, *ai* is heavy due to nucleic moras, therefore staying stressless incurs a WSP(N) violation. As a consequence, ranking WSP(N) over ALIGN $\sigma$ O generates the right result. So far then, we have achieved two rankings:


- (26) WSP >> ALIGN-HD-R [cf. (23)]  
WSP(N) >> ALIGN $\sigma$ O [cf. (25)]

Further rankings are possible. Among syllables of equal weight, the preference is to stress the one that is also onsetful, yielding ALIGN $\sigma$ O >> ALIGN-HD-R.

- (27) ALIGN $\sigma$ O >> ALIGN-HD-R  
BVV.VV: gao.ii “proper name” [E&E: 709]

	ga <sup>μ</sup> o <sup>μ</sup> i <sup>μμ</sup>	ALIGN $\sigma$ O	ALIGN-HD-R
a.	ga <sup>μ</sup> o <sup>μ</sup> .i <sup>μμ</sup>		*
b.	ga <sup>μ</sup> o <sup>μ</sup> .i <sup>μμ</sup>	*!	

This then results to the ranking:

- (28) WSP(N) >> ALIGN $\sigma$ O >> ALIGN-HD-R  
  
WSP

What this implies is that the position of WSP cannot be accurately specified. We only know that it must rank above ALIGN-HD-R. For instance, if we try to rank WSP(N) with respect to WSP, we see that no ranking argument obtains, simply because whenever WSP(N) is violated, WSP will also be violated too. As a result, a candidate that violates WSP(N) will always be worse (29b). Consequently, ranking the two constraints side-by-side yields the right results.

(29) WSP(N), WSP

BVV.PV: bi<sup>μ</sup>.si “blood” [Everett p.c.]

	bi <sup>μ</sup> .si <sup>μ</sup>	WSP(N)	WSP
a.	bi <sup>μ</sup> .s <sup>μ</sup> i <sup>μ</sup>		*
b.	bi <sup>μ</sup> .s <sup>μ</sup> i <sup>μ</sup>	*!	*

The next tableau shows what happens if we try to form a ranking argument between WSP and ALIGN<sub>σ</sub>O.

(30) WSP >> ALIGN<sub>σ</sub>O

(BV).VV.BV: boaagá “thirsty” [Everett p.c.]

	bo <sup>μ</sup> .a <sup>μ</sup> ga <sup>μ</sup>	WSP	ALIGN <sub>σ</sub> O
a.	bo <sup>μ</sup> .a <sup>μ</sup> .ga <sup>μ</sup>		*
b.	bo <sup>μ</sup> .a <sup>μ</sup> .ga <sup>μ</sup>	*!	

While this ranking generates the right result and seems to provide a ranking argument for WSP and ALIGN<sub>σ</sub>O, this is not quite the case. The reason is that we cannot produce an example that only involves a conflict between these two constraints only. Here for instance, one could use instead WSP(N) which we know is higher-ranked than ALIGN<sub>σ</sub>O. The same result would obtain. Similarly, we cannot be sure that ALIGN<sub>σ</sub>O is exclusively involved here. It could be ALIGN-HD-R instead, which would also suffer one violation. The point is that while WSP >> ALIGN<sub>σ</sub>O is compatible with the facts, it is not a necessary outcome. However, for expository reasons I will assume that WSP ranks next to WSP(N) and above ALIGN<sub>σ</sub>O. In the light of the above, the final ranking for Pirahã is shaped as shown below.

(31) *Final ranking:* WSP(N), WSP >> ALIGN<sub>σ</sub>O >> ALIGN-HD-R

One final point can be clarified. Although we have seen that among equal candidates for stress, the rightmost one actually gets it, we have not yet considered an example illustrating this point. (32) shows how the OT analysis would handle such a case. All the syllables in [k<sup>μ</sup>o<sup>μ</sup>?<sup>μ</sup>o<sup>μ</sup>p<sup>μ</sup>a<sup>μ</sup>] are bimoraic owing one of their moras to the nucleus and the other to the onset. They are thus identical in terms of their weight structure. In this environment we can see that the rightmost syllable will receive stress due to perfect right alignment.



(32) ALIGN-HD-R seen in action

PV.PV.PV: ko.ʔo.pa 'stomach' [E: 239]

	ko <sup>μ</sup> ʔo <sup>μ</sup> pa <sup>μ</sup>	WSP(N)	WSP	ALIGN <sup>σ</sup> O	ALIGN-HD-R
a.	k <sup>μ</sup> o <sup>μ</sup> .ʔ <sup>μ</sup> o <sup>μ</sup> .p <sup>μ</sup> a <sup>μ</sup>		**		**!
b.	k <sup>μ</sup> o <sup>μ</sup> .ʔ <sup>μ</sup> o <sup>μ</sup> .p <sup>μ</sup> a <sup>μ</sup>		**		*!
c.	k <sup>μ</sup> o <sup>μ</sup> .ʔ <sup>μ</sup> o <sup>μ</sup> .p <sup>μ</sup> a <sup>μ</sup>		**		

The next two tableaux merely show how the right output is chosen when all the constraints discussed are simultaneously considered.

(33) PVV.BVV.PV.VV: poo.gáí.hi.ái 'species of fruit' [E: 209]<sup>8</sup>

	po <sup>μμ</sup> ga <sup>μ</sup> i <sup>μ</sup> hi <sup>μ</sup> a <sup>μ</sup> i <sup>μ</sup>	WSP(N)	WSP	ALIGN <sup>σ</sup> O	ALIGN-HD-R
a.	...ga <sup>μ</sup> i <sup>μ</sup> .h <sup>μ</sup> i <sup>μ</sup> .a <sup>μ</sup> i <sup>μ</sup>	*	**		**
b.	...ga <sup>μ</sup> i <sup>μ</sup> .h <sup>μ</sup> i <sup>μ</sup> .a <sup>μ</sup> i <sup>μ</sup>	**	**!		*
c.	...ga <sup>μ</sup> i <sup>μ</sup> .h <sup>μ</sup> i <sup>μ</sup> .a <sup>μ</sup> i <sup>μ</sup>	*	**	*!	

In (33b), none of the two syllables that are heavy due to nucleic weight gets stress. Consequently WSP(N) is violated twice. This means one extra violation when considering the two top-ranked constraints, compared to the other candidates. (33b) is thus excluded. (33a) and (33c) incur the same number of violations with respect to WSP(N) and WSP. Since they tie, ALIGN<sup>σ</sup>O is employed to discard the candidate with the stressed onsetless syllable (33c). (33a) rightfully wins although it incurs violations of ALIGN-HD-R, but this is too low ranked to affect the outcome.

The interesting point with (34) below is that again it highlights the importance of gradient assessment of WSP (cf. (23)). (34a) which is the intended winner, can only win if it incurs one less violation of WSP by virtue of the stressed super-heavy syllable.

(34) PVV.BVV.PV: ʔaí.bái.ʔi 'much' [E: 223]

	ʔa <sup>μ</sup> i <sup>μ</sup> ba <sup>μ</sup> i <sup>μ</sup> ʔi <sup>μ</sup>	WSP(N)	WSP	ALIGN <sup>σ</sup> O	ALIGN-HD-R
a.	ʔ <sup>μ</sup> a <sup>μ</sup> i <sup>μ</sup> .bá <sup>μ</sup> i <sup>μ</sup> .ʔ <sup>μ</sup> i <sup>μ</sup>	*	**		**
b.	ʔ <sup>μ</sup> a <sup>μ</sup> i <sup>μ</sup> .bá <sup>μ</sup> i <sup>μ</sup> .ʔ <sup>μ</sup> i <sup>μ</sup>	*	**!		*
c.	ʔ <sup>μ</sup> a <sup>μ</sup> i <sup>μ</sup> .bá <sup>μ</sup> i <sup>μ</sup> .ʔ <sup>μ</sup> i <sup>μ</sup>	**	**!		

To sum up, the Pirahã weight system boils down to three different factors: a) priority of nucleic weight, b) moraicity of voiceless onsets and c) preference for onsetful stressed

<sup>8</sup> In this four-syllable word, the heaviest syllable is in fact the first *poo*. But this is beyond the stress-bearing trisyllabic window, and thus can never receive stress.

syllables. I have proposed an analysis that integrates all these points and generates the right results.

### **2.2.3 An alternative: Gordon (2005)**

In §1.3.3, we had briefly reviewed Gordon's (2005) proposal on onset weight, which was part of a larger framework to determine the weight criteria languages use. By weight criteria, Gordon means whether distinctions between: i) low vs. non-low vowels, ii) diphthongs vs. monophthongs, iii) presence of an onset vs. lack of it, iv) H-tone vs. L-tone, etc. have effects on stress attraction in favour of the first member in each of these groups. The claim is that such criteria are motivated as a function of what is both phonetically effective and phonologically simple (Gordon 1999). Phonetic effectiveness is measured by means of the perceptual energy of the rime (the algorithm to calculate this is in Gordon 2005: 606-607). However, languages only choose to use those phonetically effective weight criteria which also happen to be phonologically simple. If they are complex, they are discarded. Phonological complexity is defined along the following lines.

#### **(35) *Definition of complexity***

a) A weight distinction is complex iff it refers to more than one place predicate<sup>9</sup>.  
OR

b) It makes reference to disjunct representations of the syllable.

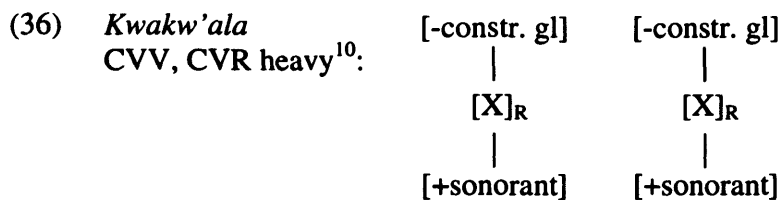
In his representations, Gordon uses a mixed system of timing slots - rather than moras - which are featurally specified. These slots are linked to place predicates, i.e. place and vowel features (Gordon 1999: 152), such as [high], [low], [back], [labial], etc. and to other non-place phonological predicates, such as root nodes, as well as manner and laryngeal features. Timing slots themselves are considered non-place phonological predicates.

To see how a language can fulfil the complexity metric, let us consider Kwakwala, which treats CVV and CVR (where R is a non-glottal sonorant) as heavy. Kwakwala fulfils both complexity criteria. First, one can refer to both types of syllables in a uniform manner, since both have two rimal root nodes where each can be described

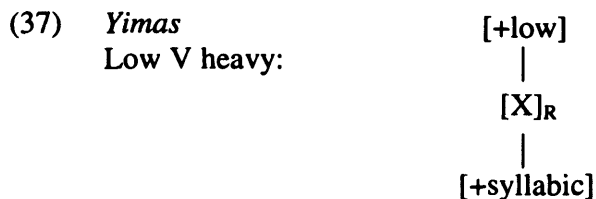
---

<sup>9</sup> This is the version as it appears in Gordon (2005: 613). In Gordon (1999: 154), the second part of the metric is the same but the first states that "a weight distinction is complex iff: i) It refers to >1 predicate in >1 dimension", so that it is not restricted to place predicates only. Independently of the definition chosen, the problem illustrated with respect to long low vowels remains.

as [-constricted glottis] and [+sonorant]. This satisfies (35b). (35a) is also satisfied because only non-place predicates need to be employed to describe the heavy syllables in this language. More specifically, we have 2 root nodes plus 2 [-const.gl.] plus 2 [+sonorant] features giving a total of 6 non-place predicates without any place predicates present. The complexity metric is satisfied.



Now consider Yimas which only treats syllables with a low vowel as heavy. Since only one syllable type can count as heavy, obviously criterion (35b) is satisfied. The question is whether (35a) is also satisfied. Note that the representation in (37) not only includes two non-place predicates (the timing slot and the [+syllabic] feature), but also one place predicate ([+low]), but this is fine since the complexity metric only penalises reference to two or more place predicates.

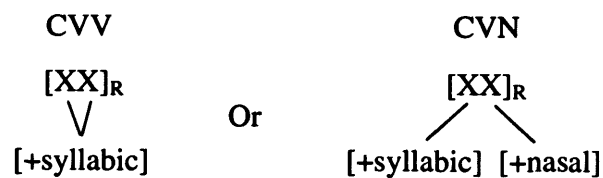


This naturally brings us to the question of which systems are then actually complex and banned by the metric in (35). One type of systems includes those which use disjunct representations for heavy syllables thus violating (35b). This means that no single syllable representation can be employed to refer to the heavy syllables of the language. One such case is a language that would consider CVV and CVN, where N=Nasal, as heavy, but CVL, where L=liquid as light. Here, there is no way that both syllable types can be subsumed under a single representation, since CVN needs an extra [+nasal] feature that CVV lacks.

<sup>10</sup> The subscript R indicates that these are rimal root nodes.

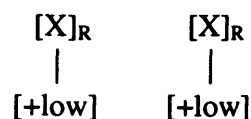
(38) *Impossible systems*

CVV, CVN heavy:



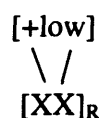
Other possible languages would have complex weight systems by making reference to more than one place predicates. Gordon claims that this would for instance occur in a hypothetical language with distinctive length for all vowels, but where only long low ones are treated as heavy<sup>11</sup>. Gordon claims that the required representation is the one below.

(39) *Long low heavy:*



Gordon (1999: 161) claims that this representation violates (35a), because it includes two non-place predicates, i.e.  $[\text{X}]_R$ , and two place predicates, namely  $[+low]$ . His model then correctly eliminates this unattested system. It should however be evident that this representation is flawed. Gordon's position, as previously mentioned, is that place predicates merely refer to place features. All standard feature theories hold that a long vowel is not specified twice for place features, but merely possesses two links between its timing positions and the (same) place feature as shown in (40). In fact, Gordon himself proposes this structure for a long low vowel, before presenting its representation according to the complexity metric (Gordon 1999: 160).

(40) *Long low V is heavy:*



The problem for Gordon now is that this structure no longer violates clause (35a) of the complexity metric, because while it still comprises two non-place predicates, it only has *one* place predicate. In this - more reasonable - understanding of a long low vowel, this weight distinction is rendered simple leaving the lack of systems with only long low Vs

<sup>11</sup> Reference to distinctive length for all vowels is important, because there are languages such as Kara (de Lacy 1997) which do treat long low vowels as heavy. Crucially however,  $[\text{a:}]$  is the only possible long vowel in that language, so Gordon claims that in this case the non-place feature  $[+syllabic]$  could be used instead of the place  $[+low]$  turning the representation into a simple one by just including non-place predicates.

as heavy unaccounted for. More generally, the choice of features and their use in Gordon is rather dubious, which should make us sceptic towards the complexity metric as a whole.

Nonetheless, I will set this problem aside and for expository reasons will assume that Gordon's approach is on the right track, so that I proceed in showing how it fails to account for Pirahã which is our focus here. Pirahã's weight system is claimed to conform to the complexity metric, since it makes use of the following simple weight distinctions.

(41) *Representations of Pirahã weight distinctions*

- |                                    |  |                                    |
|------------------------------------|--|------------------------------------|
| a. (C)V V > C V                    | b. P V(V) > B V(V)                                     | c. C V(V) > V(V)                   |
| [[X X] <sub>R</sub> ] <sub>σ</sub> | [X <sub>[-voice]</sub> [X] <sub>R</sub> ] <sub>σ</sub> | [X [X] <sub>R</sub> ] <sub>σ</sub> |

These are essentially the three components that prove crucial for Pirahã's weight system. (41a) treats syllables with long vowels as heavy irrespective of the presence of an onset, (41c) refers to the presence of an onset, while (41b) focuses on the quality of the onset. Indeed, none of these representations is complex given the definition in (35), but the false impression is given that all these representations are independent from one another and sufficient to account for the weight distinctions. The following example will show why this is not the case. Suppose we took the word *(gi)ʔaapigió* "on your arm" (the syllable *gi* is often left out; Everett p.c.). The first syllable is outside the trisyllabic window so I omit it. I will present the remaining syllables with abstract notations as in (42):

- (42)  [41a]  
[41b]

The syllable that gets stressed is the first one, because it constitutes the heaviest one. However, to reach this conclusion, we must consider the weight criteria of the language in (41) *as a whole*. To determine that PVV attracts stress more than PV we need to employ (41a), while for PVV to be a stronger contender than BVV, (41b) has to be utilized. Once these criteria are simultaneously considered, they are tantamount to disjunct representations of the syllable rendering the weight system complex and as such expected to be unattested.

More generally, if it is permitted for a language like Pirahã with larger than binary weight distinctions to be broken down into several components, as in (41a-c), then a much too powerful system is generated, where in principle a great deal of unattested systems should be able to arise.

Finally and on a more technical note, Gordon provides an OT analysis of onset-sensitive stress languages. To do that, he introduces a number of prominence constraints and provides a universal ranking for them ((43)-(44)). However, no ranking exists between the constraints in (43) and (44), thus they can be interleaved with one another so long as the universal rankings are maintained.

(43) Prom  $[X_{[-voice]}[X]_R]_\sigma \gg$  Prom  $[X_{[+voice]}[X]_R]_\sigma$  : it is more important to stress syllables with voiceless onsets than those with voiced ones.

(44) Prom  $[[XX]_R]_\sigma \gg$  Prom  $[[X]_R]_\sigma$ <sup>12</sup> : it is more important to stress binary rimes than unary rimes.

Furthermore, Gordon uses the constraint Prom  $[X_{[-voice]}[XX]_R]_\sigma$  which as he says is ‘a prominence constraint *conflating* voiceless onsets and branching rimes’ (2005: 643). One problem with this constraint is that it suspiciously looks like a description of a complex system too in a way that Gordon himself claims that is impossible: “Conspicuously absent are weight distinctions that simultaneously manipulate multiple phonological dimensions, e.g. distinctions which are sensitive to both onset voicing and vowel height, or vowel height and length” (2005: 612-613). The constraint Prom  $[X_{[-voice]}[XX]_R]_\sigma$  seems to be itself complex since it makes use of a weight distinction that simultaneously refers to onset voicing and length.

But even if we assume that this is not complex, another issue emerges. How exactly are these constraints ‘conflated’? Among the existing mechanisms, a straightforward way to achieve this result is through constraint conjunction (Smolensky 1993, Moreton and Smolensky 2002, Crowhurst and Hewitt 1997), i.e. something like Prom  $[X_{[-voice]}[X]_R]_\sigma \wedge$  Prom  $[[XX]_R]_\sigma$ . But of course if these constraints can act in conjunction they should also be able to act independently<sup>13</sup>.

<sup>12</sup> Prom  $[[X]_R]_\sigma$  is not mentioned directly in Gordon, but apart from being a very natural one in a hierarchy like (44), it can also be inferred from constraints like Prom  $[X_{[-voice]}[X]_R]_\sigma$ .

<sup>13</sup> Even if this argument were not available, still given the hierarchies in (43) and (44), these constraints would be able to interact.

Given the fixed rankings established in (43) and (44), only six total rankings are generated by the permutation of the corresponding constraints. Considering only a couple of them suffices to make the intended point.

$$(45) \quad \text{Prom } [X_{[-\text{voice}]}[X]_R]_\sigma \gg \text{Prom } [X_{[+\text{voice}]}[X]_R]_\sigma \gg \text{Prom } [[XX]_R]_\sigma \gg \text{Prom } [[X]_R]_\sigma$$

$$\text{PV} \quad > \quad \text{BV} \quad > \quad \text{VV} \quad > \quad \text{V}$$

$$(46) \quad \text{Prom } [X_{[-\text{voice}]}[X]_R]_\sigma \gg \text{Prom } [[XX]_R]_\sigma \gg \text{Prom } [X_{[+\text{voice}]}[X]_R]_\sigma \gg \text{Prom } [[X]_R]_\sigma$$

$$\text{PV} \quad > \quad \text{VV} \quad > \quad \text{BV} \quad > \quad \text{V}$$

Both these systems are somehow reminiscent of Pirahã, but not quite, since they produce systems where  $\text{PV} > \text{BV} > \text{VV}$  or  $\text{PV} > \text{VV} > \text{BV}$  - which to my knowledge are unattested - instead of  $\text{VV} > \text{PV} > \text{BV}$ . This can be produced by a ranking like:

$$(47) \quad \text{Prom } [[XX]_R]_\sigma \gg \text{Prom } [X_{[-\text{voice}]}[X]_R]_\sigma \gg \text{Prom } [X_{[+\text{voice}]}[X]_R]_\sigma$$

$$\text{VV} \quad > \quad \text{PV} \quad > \quad \text{BV}$$

The role of  $\text{Prom } [[X]_R]_\sigma$  is not of particular importance since monovocalic onsetless syllables are banned from Pirahã. But even in this system (47), constraint conjunction would not be avoided and additional constraints like Gordon's  $\text{Prom } [X_{[-\text{voice}]}[XX]_R]_\sigma$  are in need to produce the  $\text{PVV} > \text{BVV}$  part of the scale. In sum, Gordon's formalisation greatly over-generates, while it has inherent problems with the conjunctions and constraints it allows.

Finally, recall from (41), repeated here as (48), that three criteria intermingle to produce the Pirahã weight hierarchy.

(48) *Representations of Pirahã weight distinctions*

a. $(C)\text{VV} > \text{CV}$ $[[X \ X]_R]_\sigma$	b. $\text{PV}(V) > \text{BV}(V)$ $[X_{[-\text{voice}]} [X]_R]_\sigma$	c. $\text{CV}(V) > \text{V}(V)$ $[X [X]_R]_\sigma$
---	--	---

Setting aside issues discussed before concerning the complexity of the system, an additional consideration emerges. Gordon's complex constraints seem to miss the basic insights and what's more duplicate information. Thus for example, he uses both  $\text{Prom } [X_{[-\text{voice}]}[X]_R]_\sigma$  and the dubiously conflated  $\text{Prom } [X_{[-\text{voice}]}[XX]_R]_\sigma$ , implying that they are two different things, while the strong intuition is that what only needs to be expressed here is the influence a voiceless onset has on the total syllable weight. In sum,

vague mechanisms like ‘conflation’ as well as complex constraints which superficially have different statements weaken Gordon’s system and make its implementation implausible.

In this section we have explored Gordon’s theory of weight (Gordon 1999) and its application to onset-sensitivity (Gordon 2005). Certainly a number of interesting findings arise, such as the role of perceptual energy and its relation to the auditory boost that an onset can offer to the vowel following it (see §1.3.3) as well as a more unified way of modeling onset-sensitive stress languages compared to previous analyses. Despite this, a number of flaws are evident which seriously undermine the theory as a whole and render it untenable. These include both phonetic and phonological considerations. The former, discussed in §1.3.3, involve the use of perceptual energy only through its relation with the nucleic vowel disregarding the perceptual energy of the onset itself. The latter, discussed in the present section, refer to weaknesses and incorrect predictions of the complexity metric as well as the problematic use of prominence constraints. Moreover, certain difficulties arise with the formalisation of the theory that involve peculiar and unexplained mechanisms such as conflation and the loss of insight through duplication of constraints.

## **2.3 Arabela**

### **2.3.1 Data and generalisations**

The next language I wish to discuss is Arabela, a Zaparoan language spoken in Peru by about 50 Indians (Rich 1963, Payne and Rich 1988). Arabela is very interesting because although it generally presents a normal rhythmic stress algorithm, this pattern is disrupted in a particular context that has to do with the voicing of the onsets involved. My stress examples all come from Payne and Rich (1988). A preliminary description of stress appears in Rich (1963), which is superseded by Payne and Rich where the effects of onsets had been noticed in a published form for the first time.

I will start with the basics of the language by first presenting the syllable structure and working my way to the segmental inventory, since the latter is going to prove relevant for the stress facts we examine here. Syllable structure is mentioned in Rich (1963), but several points remain dubious. It is claimed that onsetless syllables and codas are allowed, but examples are usually ambiguous. For instance, words such as /rupaa/ ‘mouth’ or /siinu/ ‘to raise a creature’ are syllabified according to Rich (1963:



201) as [ru.pa.a] and [si.i.ni]. It is also noted that in such V syllables, as they are referred to, the V is always the same as the preceding one. This makes it quite likely that this is actually a long vowel we are talking about.

As for codas, to the extent that these exist, they are most of the time glides as in [kway.ni.nyu] ‘to hunt’ or [su.wo.kwaw] ‘ground corn’, although other sequences that seem to involve sonorant codas also arise, e.g. /sapartu/ ‘shoulder blade’ or /mante/ ‘moth’. Other work however presents Arabela as a representative example of a language that only has CV and CCV syllables (Levelt and van de Vijver 1998, Gussenhoven and Jacobs 2005). This conflict is not resolved in the more recent of the Arabela papers, i.e. Payne and Rich (1988), where no example includes anything that can be construed as a coda or as long vowel. Importantly though, even if codas do exist in the language, they are irrelevant for weight and stress purposes, e.g. (mòkof)(tyáka) ‘palm fruit’ instead of \*(mò)(kòf)(tyáka), which means that codas are not moraic. In the absence of convincing evidence, I will follow Levelt and van de Vijver (1998) and Gussenhoven and Jacobs (2005) in assuming that the language only allows CV and CCV syllables<sup>14</sup>.

We can now move to the phonemic inventory, which is presented in (49) and followed by some examples in (50).

(49) *Arabela phonemic inventory*

vowels: *i, e, a, o, u*

consonants: stops *p, t, k*, fricatives *s, f, h*, nasals *m, n*, liquid *r* and glides *w, y*

(50) a. stops

<i>pinyu</i>	“to hit”	<i>tinyu</i>	“to fall”
--------------	----------	--------------	-----------

<i>kinyu</i>	“to stay”		
--------------	-----------	--	--

b. fricatives

<i>siyokwa</i>	“tucuayo bird”	<i>fiyokwa</i>	“grease”
----------------	----------------	----------------	----------

c. nasals

<i>miyano</i>	“plaything”	<i>niyano</i>	“he is coming”
---------------	-------------	---------------	----------------

d. liquid – glides

<i>riyano</i>	“he is breathing”	<i>tawe</i>	“foreigner”
---------------	-------------------	-------------	-------------

<i>hayunu</i>	“pulling”		
---------------	-----------	--	--

While the language seems to lack voiced obstruents phonemically, these arise as variants of the voiceless stops in certain contexts which are unfortunately not entirely

<sup>14</sup> This is why we cannot confidently group Arabela either along languages like Pirahã where the presence of the onset matters or languages like Karo where it does not. If Arabela lacks onsetless syllables altogether, its behaviour in this respect cannot be tested for obvious reasons.

clear given Rich's (1963) discussion<sup>15</sup>, but can be seen in (51c-d) below. Note especially that in the morpheme meaning "father" there is alternation between k~x~g.

(51)	/ke/	[kiʔ]	"father"
	/nake/	[nāxiʔ]	"his father"
	/kanaake/	[kanāāgiʔ]	"our (excl.) father"
	/saako/	[saayoʔ]	"corn"

As a result, when I refer to 'voiced onsets' below, I will actually refer to sonorant onsets. As we will see in a moment, Arabela possesses a process of stress shift caused by voiced onsets in a specific environment. Although all the cited examples involve sonorants, the use of [voice] as the relevant feature rather than [son] is justified by the fact that the authors of the original source, i.e. Payne and Rich (1988), talk about voicing. Had the authors intended reference to sonorants only, I believe they would have been explicit about sonorancy instead of voicing since they must have been aware of the voiced obstruents mentioned in Rich (1963) (given that Rich is one of the authors of Payne and Rich 1988). This finding also fits nicely with the general picture drawn here, therefore I will assume that Arabela sonorants are specified as [voice] and anticipate that voiced obstruents would pattern like the sonorants do.

As already mentioned, stress is rhythmic and creates trochees from left to right. The rightmost stress is the primary one (52). Degenerate monosyllabic feet are also admitted as (52b-c) reveal.

(52)	a. tēnakári	'afternoon'
	b. sàmarú	'spirit'
	c. hùwahàniyá	'peaceful'

The interesting exceptional pattern is illustrated below. According to Payne and Rich (1988), "if a word-final syllable that would have received stress has a voiced onset, and the immediately preceding syllable has a voiceless onset, then the syllable with the voiceless onset is stressed".

(53)	a. nòwafɿfáno	*nòwafɿfanó	'brightened'
	b. sàpohòsáno	*sàpohòsanó	'deceived'
	c. mwèratityénu	*mwèratityenú	'cause to be seen'

<sup>15</sup> To be more precise, Rich (1963) suggests that the alternants occur freely and progressively lenited as one moves further within the phrase. But this is not supported by the more recent description of Payne and Rich (1988) who treat the obstruents involved as voiceless stops only.

In these examples we would normally expect stress on the first, third and fifth syllables, but instead stress docks on the first, third and *fourth* syllable. My proposal is that this pattern can be explained by making use of moraic onsets in a manner similar to Pirahã. Voiceless onsets are moraic, voiced ones are not. The Weight-to-Stress Principle is active in this language too and requires heavy, i.e. PV, syllables to receive stress. Anticipating the detailed analysis that follows, we can at this stage state that this property, combined with the fact that feet in Arabela are rhythmic and prefer to align with the right edge of the word, will give us the explanation why such stress shift occurs only when the penult has a voiceless onset and the ultima a voiced one.

Recall that the voiced consonants in Arabela are actually the sonorant ones. Thus, despite the apparent lack of overt presence of voiced obstruents, I would like to argue that the moraicity of Arabela onsets is identical to that of Pirahã. Consonants lacking [voice] attract stress more than their [voice] counterparts by virtue of their moraicity. The onset moraicity ranking in (54) captures this point provided sonorants are specified as [voice], as we have argued throughout. *Thus, for Arabela, I will henceforth interchangeably use the words voiced consonants and sonorants to refer to a single entity, namely that of [voice] consonants. I will also occasionally characterize sonorants as voiced to remind the reader that I assume that they possess the [voice] feature.*

(54) *Onset moraicity in Arabela*

$*\mu/\text{ONS}/[\text{voice}] \gg \text{MORAIC} \gg *\mu/\text{ONS}$

As (54) suggests, the distribution of moraic onsets in Arabela is identical to that of Pirahã, in that only voiceless onsets are moraic, whereas voiced sonorants are not. Note however, that unlike Pirahã where moras of both nuclei and onsets were presented, in Arabela, to facilitate the reading of tableaux, I only refer to onset moras since there are no examples that involve long vowels (whose existence is generally questionable).

(55)  $*\mu/\text{ONS}/[\text{voice}] \gg \text{MORAIC} \gg *\mu/\text{ONS}$

	pohonu	$*\mu/\text{ONS}/[\text{voice}]$	MORAIC	$*\mu/\text{ONS}$
a.	p <sup>h</sup> oh <sup>h</sup> on <sup>h</sup> u	*!		***
b.	pohonu		***!	
c.	poh <sup>h</sup> onu		**!	*
d.	p <sup>h</sup> oh <sup>h</sup> onu		*	**

In this (hypothetical, based on the actual /rupohonu/) word, the first candidate assigns a mora to a sonorant causing violation of the high-ranked constraint which militates against voiced moraic onsets, thus it is ruled out. From the remaining candidates, only the last survives because it satisfies MORAIC to the best extent possible; the other two fail to assign moras to all or one of the voiceless consonants incurring serious violations of MORAIC. All in all, this tableau evinces that moraic onsets in Arabela are the voiceless ones, whereas the voiced are not.

Now, we have just talked about the moraicity of the onsets, without having shown their effect on Arabela stress. This point is illustrated below, where the constraints introduced so far, make no claims about the footing of the words. Any of (56a) or (56b) could be the winner. Numerous other ingredients are required to achieve the right results. These are explored in the next section.

(56) \* $\mu$ /ONS/[voice] >> MORAIC >> \* $\mu$ /ONS

	pohonu	* $\mu$ /ONS/[voice]	MORAIC	* $\mu$ /ONS
☞ a.	(p <sup>h</sup> oh <sup>h</sup> o)(nu)		*	**
☞ b.	(p <sup>h</sup> o)(h <sup>h</sup> onu)		*	**

## 2.3.2 Analysis of the stress system

### 2.3.2.1 The basic analysis

Recall from (52) [repeated here as (58)] that the default stress system of Arabela involves trochees rhythmically formed from left to right. It is also the case that the rightmost stress is the primary one, therefore the following two constraints are undominated in Arabela. Since these are never violated, I will not consider them in the tableaux that follow.

(57) TROCHEE: Feet have initial prominence (Kager 1999)

ALIGN-HDFT-R: Align-R (HdFt, PrWd)

The head foot is rightmost in the prosodic word (Hyde 2001)

(58) *Rhythmic stress in Arabela*

- a. tènakári                    ‘afternoon’
- b. sàmarú                    ‘spirit’
- c. hùwahàniyá            ‘peaceful’

Additional prosodic constraints are needed to account for the parsing of all syllables into feet, the directionality of footing and the size of feet. These are listed next.

(59) PARSE- $\sigma$ : Syllables are parsed by feet (Prince and Smolensky 1993/2004)

FTBIN(MAX)<sup>16</sup>: Feet are maximally bisyllabic<sup>17</sup> (Everett 1996)

ALL-FT-R: Align-R (Ft, PrWd) (McCarthy and Prince 1993)

Align the right edge of every foot with the right edge of the PrWd

ALL-FT-L: Align-L (Ft, PrWd) (McCarthy and Prince 1993)

Align the left edge of every foot with the left edge of the PrWd

We will take each of these in turn and see how they are ranked in Arabela. The first point we can easily show is that the rightward directionality of footing establishes that ALL-FT-R >> ALL-FT-L.

(60) ALL-FT-R >> ALL-FT-L

	samaru	ALL-FT-R	ALL-FT-L
a.	(sàma)(rú)	*	**
b.	(sà)(máru)	**!	*

We can now examine the relationship between syllable parsing and foot alignment. FTBIN(MAX) needs to dominate ALL-FT-R as illustrated below.

(61) FTBIN(MAX) >> ALL-FT-R, cf. [hùwahàniyá] (58c)

	σσσσσ	FTBIN(MAX)	ALL-FT-R
a.	(σσσσσ)	*!	
b.	(σσσ)(σσ)	*!	**
c.	(σσ)(σσσ)	*!	***
d.	(σσ)(σσ)(σ)		****

<sup>16</sup> The more traditional FTBIN runs into troubles. PARSE- $\sigma$  would need to dominate FTBIN since all syllables are parsed into feet. ALL-FT-R is needed to account for the rightward directionality of footing. The problem arises with the interaction of FTBIN and ALL-FT-R as observed in large odd-syllable words where either too many syllables could be stuffed into a foot or ternary feet are preferred. None of these actually occurs. The problem is resolved once we use high-ranking FTBIN(MAX) to allow feet to be unary or bisyllabic. In this case FTBIN(MIN), i.e. feet are minimally bisyllabic, must be assumed to be very low-ranked since degenerate feet are admitted.

<sup>17</sup> FTBIN requires feet to be binary either on the syllabic or moraic level. In Arabela we claim that FTBIN holds on the syllabic level, thus feet have to be bisyllabic. This is interesting since the language is claimed to be weight sensitive, but is analysed as quantity-insensitive when it comes to foot formation. Perhaps a similar claim can be made for iambic feet in Paumari. Based on the fact that in this language long vowels can surface in the non-head position of an iamb, Everett (2003) argues that they are monomoraic. It could be possible though to claim that long vowels are bimoraic, but that the language forms feet on the syllabic rather than the moraic level. This is quite strange given that generally iambs favour durational differences (cf. the Iambic/Trochaic Law in Hayes 1995). However, Everett shows convincingly that Paumari does not seem to adhere to the law for independent reasons, thus making the QI formation of feet less peculiar.

The first three candidates fail FTBIN(MAX) as they include either ternary or even larger feet, which exceed the bisyllabic maximum. The winning candidate is (d), since its unary and bisyllabic feet perfectly satisfy FTBIN(MAX) even at the expense of ALL-Ft-R. We can also easily show that PARSE- $\sigma$  >> ALL-Ft-R.

(62) PARSE- $\sigma$  >> ALL-Ft-R

	tenakari	PARSE- $\sigma$	ALL-Ft-R
☞	a. (tèna)(kári)		**
	b. tena(kári)	**!	

It remains to see what the relationship between PARSE- $\sigma$  and FTBIN(MAX) is. In fact, these constraints do not conflict with one another, so we can assume that they are placed next to each other (63).

(63) PARSE- $\sigma$ , FTBIN(MAX)

	samaru	PARSE- $\sigma$	FTBIN(MAX)
☞	a. (sàma)(rú)		
	b. (sáma)ru	*!	

The prosodic ranking we have thus far established is presented and illustrated in (64).

(64) PARSE- $\sigma$ , FTBIN(MAX) >> ALL-Ft-R

	tenakari	PARSE- $\sigma$	FTBIN(MAX)	ALL-Ft-R
☞	a. (tèna)(kári)			**
	b. (ténakari)		*!	

To sum up then, so far we have achieved two rankings as illustrated below.

- (65) a.  $*\mu/\text{ONS}/[\text{voice}]$  >> MORAIC >>  $*\mu/\text{ONS}$   
 b. PARSE- $\sigma$ , FTBIN(MAX) >> ALL-Ft-R >> ALL-Ft-L

(65a) tells us which onsets may surface with moras in the output, whereas (65b) illustrates the general stress pattern of the language. These rankings though appear unrelated and seem to make statements regarding different aspects of the language. There is nonetheless a link between them which I would like to suggest is the WSP.

(66) WSP (Weight-to-Stress Principle): Heavy syllables must be stressed

Furthermore, I propose that the contextual stress shift depends on the interaction between WSP and ALL-Ft-R. This makes it imperative to locate where WSP is positioned in each of the rankings in (65).

First consider what effect the inclusion of WSP in the moraic markedness hierarchy (65a) has. It can actually single out just one winner for the problematic case we had considered in (56), where no footing could be determined at that stage. Note that right now there is no ranking argument yet about the exact location of WSP. I will assume that it is placed next to MORAIC, but will return to this issue in section §2.3.4.

(67) \* $\mu$ /ONS/[voice] >> MORAIC, WSP >> \* $\mu$ /ONS

	pohonu	* $\mu$ /ONS/[voice]	MORAIC	WSP	* $\mu$ /ONS
a.	(p <sup>μ</sup> òh <sup>μ</sup> o)(nú)		*	*!	* *
b.	(p <sup>μ</sup> ò)(h <sup>μ</sup> ónu)		*		* *

We also need to position WSP somewhere in the ranking of (65b). To do so, we have to consider a case where WSP is active and is in conflict with any of the constraints of (65b) repeated below.

(68) PARSE- $\sigma$ , FTBIN(MAX) >> ALL-Ft-R >> ALL-Ft-L

A relevant example appears in those cases where stress retracts from the ultima to the penult as a result of the voiceless onset this includes. The word *nòwafíáno* fits this profile. Below I only consider the most appealing candidates, which satisfy the high-ranking PARSE- $\sigma$  and FTBIN(MAX) constraints.

(69) ALL-Ft-R >> ALL-Ft-L<sup>18</sup>

		ALL-Ft-R	ALL-Ft-L
☛	a. (nòwa)(ʃ <sup>μ</sup> íʃ <sup>μ</sup> a)(nó)	**** (4)	***** (6)
	b. (nò)(wàʃ <sup>μ</sup> i)(ʃ <sup>μ</sup> áno)	*****! (6)	**** (4)
☛	c. (nòwa)(ʃ <sup>μ</sup> i)(ʃ <sup>μ</sup> áno)	*****! (5)	***** (5)

<sup>18</sup> ☛ denotes a candidate which is wrongly chosen as a winner.

This evaluation wrongly picks out (69a) as the winning candidate, which is of course the candidate one would expect if the parsing had proceeded as usual. Taking WSP into consideration generates the correct result.

(70) ALL-Ft-R, WSP >> ALL-Ft-L

	ALL-Ft-R	WSP	ALL-Ft-L
a. (nòwa)(ʃ <sup>u</sup> <sub>1</sub> ʃ <sup>u</sup> <sub>2</sub> a)(nó)	**** (4)	*	*****! (6)
b. (nò)(wàʃ <sup>u</sup> <sub>1</sub> i)(ʃ <sup>u</sup> <sub>2</sub> áno)	*****! (6)	*	**** (4)
c. (nòwa)(ʃ <sup>u</sup> <sub>1</sub> i)(ʃ <sup>u</sup> <sub>2</sub> áno)	***** (5)		***** (5)

In (70), candidate (b) loses in the very beginning, but both (a) and (c) survive since they incur the same number of total violations with respect to the unranked ALL-Ft-R and WSP. In particular, although (c) fares worse than (a) in terms of right alignment, it totally satisfies WSP since it manages to stress both syllables with voiceless onsets. On the contrary, (a) only stresses one of the two. Thus, the competition moves on to ALL-Ft-L, which correctly selects (c) as the winning candidate.

Stress shift then is the product of a balancing effect exerted by WSP on the violations of right alignment (compare (70a) with (70c)). It is only in the case of odd-syllable words containing a sequence of voiceless-sonorant in the penult and ultima respectively, where moving stress away from the ultima and thus causing one extra ALL-Ft-R violation for the sake of better satisfaction of WSP can be justified. The total violations incurred though are in each case the same, so the decision falls on low-ranked ALL-Ft-L, which favours the stress shifted candidate, as it produces better left alignment (cf. (70c)). As we will see in a later section, stress shift in any other combination of onsets and number of syllables leads to massive violations of ALL-Ft-R which cannot be compensated for by better satisfaction of WSP. As a result, stress shift does not apply.

### 2.3.2.2 The importance of equal ranking<sup>19</sup>

The example in (70) also addresses an important theoretical issue, which merits some discussion at this point. In (70) we have already seen that ALL-Ft-R and WSP are located in the same position in the ranking. This however should not be construed as an undetermined ranking. Instead, it is crucial that these constraints are *equally* ranked.

<sup>19</sup> For a detailed discussion of equal ranking and other cases of crucially non-ranked constraints see Topintzi (2005a) and Rice (in press).



Had they been unranked with respect to one another, then we would expect that they would yield the same (correct) result whether we had assumed a ranking ALL-Ft-R >> WSP or WSP >> ALL-Ft-R. This is not the case though. If for instance we adopt ALL-Ft-R >> WSP, then in the case of stress shift, we predict instead the rhythmic pattern (cf. (70a)), while in the opposite ranking of WSP >> ALL-Ft-R, there are cases where we predict stress shift instead of rhythmic stress (cf. (72b)). These are illustrated below. The relevant candidates of (70) are repeated here in (71), while (72) follows<sup>20</sup>.

(71) i) ALL-Ft-R >> WSP >> ALL-Ft-L ---- wrong winner

		ALL-Ft-R	WSP	ALL-Ft-L
☀	a. (nòwa)(fɪfa)(nó)	**** (4)	*	***** (6)
☞	b. (nòwa)(fɪ)(fáno)	*****! (5)		***** (5)

Under the undetermined ranking approach, we would expect both ALL-Ft-R >> WSP and WSP >> ALL-Ft-R to produce the same result. Although the latter ranking produces the desirable outcome, the former fatally chooses the wrong candidate (71.i). This is not the case if the constraints are equally ranked and simultaneously evaluated at the same position as the following tableau shows.

ii) equal ranking: ALL-Ft-R, WSP >> ALL-Ft-L ---- correct winner

		ALL-Ft-R	WSP	ALL-Ft-L
	a. (nòwa)(fɪfa)(nó)	**** (4)	*	*****! (6)
☞	b. (nòwa)(fɪ)(fáno)	***** (5)		***** (5)

In the following case, the same phenomenon is illustrated by using an example where rhythmic stress occurs. This example is discussed in more detail in (82) below. This time it is evident that ALL-Ft-R >> WSP could work fine, but the reverse WSP >> ALL-Ft-R fails to generate the right outcome (72.i).

(72) i) WSP >> ALL-Ft-R >> ALL-Ft-L ---- wrong winner

		WSP	ALL-Ft-R	ALL-Ft-L
☞	a. (kòko)(náka)	**!	** (2)	** (2)
☀	b. (kò)(kòna)(ká)		**** (4)	**** (4)

<sup>20</sup> Since it has been established that \*μ/ONS/[voice] >> MORAIC >> \*μ/ONS, and to simplify representations as much as possible, from now on I will no longer indicate the moras on voiceless onsets (unless required for illustration purposes), but will simply assume them.

Again equal ranking circumvents this problem by assigning the same number of violations in both candidates and letting ALL-Ft-L determine the optimal candidate (72.ii).

ii) equal ranking: ALL-Ft-R, WSP >> ALL-Ft-L ---- correct winner

		ALL-Ft-R	WSP	ALL-Ft-L
☞	a. (kòko)(náka)	** (2)	**	** (2)
	b. (kò)(kòna)(ká)	**** (4)		****! (4)

I conclude that this property of rankings must be recognized and depicted differently from the comma used for cases of undetermined ranking. I will be using the symbol of equality “=” and the wavy line to indicate this in tableaux. Thus, the ranking of WSP, ALL-Ft-R is now updated to:

(73) EQUAL RANKING: WSP = ALL-Ft-R >> ALL-Ft-L

Effectively, WSP and ALL-Ft-R have to be equally ranked<sup>21</sup>. This leaves us with:

<sup>21</sup> Brian Jose (p.c.) suggests that no need of equal ranking is required if we assume the strict domination ranking ALL-Ft-R >> WSP >> ALL-Ft-L and also assume that the alignment constraints work categorically (McCarthy 2003b), i.e. every misaligned foot incurs just one violation of alignment ignoring the number of syllables away from the prescribed edge. While this works for some cases, it also produces variation which is unattested. The following tableau exemplifies.

(i) unattested variation due to ALL-Ft-R >> WSP >> ALL-Ft-L [with categorical alignment]

		ALL-Ft-R	WSP	ALL-Ft-L
☞	a. (s <sup>h</sup> àk <sup>h</sup> a)(màna)(h <sup>h</sup> á)	(saka) (mana)	*	(mana) (ha)
	b. (s <sup>h</sup> à)(k <sup>h</sup> àma)(na)(h <sup>h</sup> á)	(sa) (kama) (na)!		(kama) (na) (ha)
☞	c. (s <sup>h</sup> à)(k <sup>h</sup> àma)(náh <sup>h</sup> a)	(sa) (kama)	*	(kama) (naha)

In the alignment cells, one can see the misaligned feet in brackets. Each foot amounts to one violation. The second candidate has three ALL-Ft-R violations, thus it loses. (a) and (c) though fare equally well and thus are both predicted as correct winners. The data though show us that only (a) is attested. One can stretch the argument further; (c)'s violation of ALL-Ft-R may be more severe because it is two disyllabic feet that violate it, compared to one monosyllabic and one disyllabic in (a). This would rule out (c) in favour of (a). First, this re-introduces gradiency into the system, so it runs counter to the spirit of the whole re-analysis. More importantly, there are other examples such as /huwahaniya/ for which this analysis generates two outcomes too, i.e. (h<sup>h</sup>ùwa)(h<sup>h</sup>àni)(yá) and (h<sup>h</sup>ùwa)(h<sup>h</sup>à)(níya). Only the former is correct. Nonetheless, if one were to make use of the same argument, then the winner violates ALL-Ft-R twice i.e. (huwa) and (hani) with two disyllabic feet, whereas the wrong candidate fares better with (huwa) and (ha). Had this logic been valid, then (hùwa)(hà)(níya) should instead be right, which is of course incorrect. Therefore, this reanalysis does not take us too far.

(74) Mini-grammar of Arabela - First attempt

\*μ/ONS/[voice] >>      MORAIC,      >> \*μ/ONS  
WSP =  
PARSE-σ, FtBIN(MAX) >>      ALL-Ft-R >> ALL-Ft-L

One could however argue that only the PARSE-σ, FtBIN(MAX) >> ALL-Ft-R = WSP >> ALL-Ft-L part of the ranking above can be established. The relationship of WSP with the other branch of the ranking could be different. For instance, WSP could dominate all the constraints in the first line of (74) yielding:

(75) PARSE-σ, FtBIN(MAX) >> WSP = ALL-Ft-R >> ALL-Ft-L, \*μ/ONS/[voice] >> MORAIC >> \*μ/ONS

This cannot occur though. The following tableau explains why. Suppose we use another example, but equivalent in structure to the one in (70), e.g. *sàpohòsáno* “deceived”. Let us consider four candidates in total: one that has rhythmic stress; one that has stress retraction and the moraic counterparts for each of these two. Our focus is on the constraints below.

(76) Unacceptability of ALL-Ft-R = WSP >> MORAIC

		ALL-Ft-R	WSP	MORAIC
a.	(sàpo)(hò)(sáno)	*****! (5)		***** (5)
b.	(sàpo)(hòsa)(nó)	**** (4)		***** (5)
c.	(s <sup>μ</sup> àp <sup>μ</sup> o)(h <sup>μ</sup> ò)(s <sup>μ</sup> áno)	*****! (5)	*	* (1)
d.	(s <sup>μ</sup> àp <sup>μ</sup> o)(h <sup>μ</sup> òs <sup>μ</sup> a)(nó)	**** (4)	*!*	* (1)

The problem here is that by having MORAIC low-ranked, there is no reason why the rhythmic non-moraic candidate (b) is not the winner, given that it vacuously satisfies WSP and fares better than the stress-retracted contenders (a) and (c) in terms of foot alignment. Evidently this is wrong, because for this word, stress retraction occurs. If on the other hand, as in (77), MORAIC is promoted and ranked at least as highly as ALL-Ft-R - where WSP is - then (b) can no longer survive<sup>22</sup>. The competition necessarily refers to the moraic contenders (c) and (d). The role of WSP becomes now evident. Thanks to it, the violations of (d) are equal to those of (c) and thus permit both to continue to

<sup>22</sup> The exact ranking of MORAIC with respect to ALL-Ft-R and WSP is examined below in section §2.3.4.

another round of evaluation. The next constraint - ALL-Ft-L - determines the final outcome. It chooses (c) as it incurs one less violation when compared with (d)<sup>23</sup>.

(77) MORAIC, ALL-Ft-R = WSP >> ALL-Ft-L

	MORAIC	ALL-Ft-R	WSP	ALL-Ft-L
a. (sàpo)(hò)(sáno)	**** (4)	***!*** (5)		***** (5)
b. (sàpo)(hòsa)(nó)	**** (4)	***!* (4)		***** (6)
c. (s <sup>μ</sup> àp <sup>μ</sup> o)(h <sup>μ</sup> ò)(s <sup>μ</sup> áno)	* (1)	***** (5)	* (1)	***** (5)
d. (s <sup>μ</sup> àp <sup>μ</sup> o)(h <sup>μ</sup> òs <sup>μ</sup> a)(nó)	* (1)	**** (4)	** (2)	*****! (6)

What this means is that the stress shift effects are attributed to a combination of factors; MORAIC favors candidates with moraic voiceless onsets, while WSP and ALL-Ft-R ensure that the candidate with stress shift has the same number of total violations as its non-shift counterpart has. As a result, both make it to the next stage of evaluation. There ALL-Ft-L selects the candidate with stress shift, since this is the one that satisfies left-alignment more satisfactorily<sup>24</sup>.

### 2.3.2.3 Testing the system: voiceless-sonorant combinations and non-shift of stress in other positions

To recap, in Arabela stress is rhythmic. Bisyllabic feet are created wherever possible. This means that words with an even-number of syllables are fully parsed in a very ordinary binary way. It is only in odd-syllable words and in fact in a subset of these where the normal algorithm is obscured. There, the combination of relatively high-ranked MORAIC and WSP leads to the sole survival of candidates with moraic onsets. A balancing effect is achieved among the violations of WSP and ALL-Ft-R, which makes the normal rhythmic candidate and the stress shifted candidate tie. Subsequently, foot-alignment to the left edge of the word selects the candidate that satisfies it the best, i.e. the one with stress retraction. Importantly, this effect arises only in the final two

<sup>23</sup> One may wonder why (a) - the non-moraic stress retracting candidate - can never win. The reason is that if MORAIC is highly-ranked (as argued here), it will cause too many violations of it. If MORAIC is low-ranked, then (b) should still be better as it harmonically bounds (a). This is indirect evidence suggesting that moraicity must be involved in cases like the one here, otherwise stress shift cannot be explained by using more standard prosodic and alignment constraints.

<sup>24</sup> Importantly note that up to now, we have considered candidates where *all* obstruent onsets surface as moraic. This is not necessarily the case though. In §2.3.4 we explore the possibility that only the onsets of stressed syllables are moraic. As we will see, depending on the exact specifics of the ranking between WSP and MORAIC we can generate both possibilities.

syllables where primary stress appears. Below, we will explore the relevant cases in more detail and see how the system we have proposed accounts for them.

This section centres around two issues; first, whether the correct predictions are made for all possible combinations of voiced sonorants and voiceless consonants in the two last syllables of the word and second whether the appearance of a voiceless onset in a syllable other than the final two disrupts the normal algorithm. According to the description of the language (Payne and Rich 1988), it is only in words of odd-syllables where the ultima has a voiced onset and the penult has a voiceless one that stress appears on the penult rather than the anticipated ultima. Our task is to examine what happens in the other possible combinations of voiced sonorants and voiceless in the two last syllables of the word. Can the empirical data be accounted for by the mini-grammar we have offered in (74)? In practice, the following cases have to be investigated:

(78) *Full range of possible sonorant-voiceless onsets in the final two syllables*

	<i>Penult</i>	<i>Ultima</i>	<i>Example</i>	<i>Result</i>
a.	(Voiced) Sonorant	(Voiced) Sonorant	hùwahàniyá	no shift to penult
b.	(Voiced) Sonorant	Voiceless	sàkamànahá	no shift to penult
c.	Voiceless	Voiceless	kòkotáka	no shift to penult
d.	Voiceless	(Voiced) Sonorant	nòwaʃjáno	<b>shift to penult</b>

As shown before, it is the case that only in (78d) there is shift of stress. It remains to see whether these results are generated given the system proposed above. Let us first examine an example where both final syllables contain a sonorant onset.

(79) sonorant - sonorant

PARSE- $\sigma$ , FtBIN(MAX) >> ALL-Ft-R = WSP >> ALL-Ft-L

	PARSE - $\sigma$	FtBIN (MAX)	ALL-Ft-R	WSP	ALL-Ft-L
a. (hùwa)(hàni)(yá)			**** (4)		***** (6)
b. (hùwa)(hà)(níya)			*****! (5)		***** (5)
c. (hùwa)(háni)ya	*!		**** (4)		** (2)

Here WSP plays no role since all candidates satisfy it by stressing the syllables with voiceless onsets. Although, candidate (c) produces the fewest violations of the lower-ranked constraints, it leaves a syllable unparsed causing a fatal violation of PARSE- $\sigma$ . Therefore any candidates violating this constraint will never survive. (79b) is identical

to those outputs where stress indeed shifts to the penult, but here it produces an additional unnecessary violation of ALL-Ft-R, leaving (a) emerge as the winner.

(80) sonorant - voiceless

ALL-Ft-R = WSP >> ALL-Ft-L

		ALL-Ft-R	WSP	ALL-Ft-L
☞	a. (sàka)(màna)(há)	**** (4)	*	***** (6)
	b. (sàka)(mà)(náha)	***** (5)	*!*	***** (5)

In (80) there is no trigger for stress shift, since the ultima, which contains a voiceless onset, would normally receive primary stress due to rhythmic stress. Thus, (a) is unsurprisingly chosen. The final case we need to consider is one that involves voiceless onsets on both the final two syllables.

(81) voiceless - voiceless --- and no rightward stress shift

ALL-Ft-R = WSP >> ALL-Ft-L

		ALL-Ft-R	WSP	ALL-Ft-L
☞	a. (kòko)(táka)	** (2)	**	** (2)
	b. (kò)(kò)(tà)(ká)	*****!* (6)		***** (6)
	c. (kòko)(tà)(ká)	*** (3)	*	*****! (5)

This tableau is particularly informative, since not only it produces the desired result, it also shows that WSP although important, does not force the parse to restart whenever a voiceless onset is encountered [as in (b)]; instead WSP interacts with ALL-Ft-R, whose violations are decisive<sup>25</sup>. The other two rivals tie up to a certain point making it very likely that (a) loses. ALL-Ft-L saves the situation by preferring (81a) due to its fewer violations.

The example above has an additional advantage; it is an instance where rightward stress shift (cf. (81c)) could be expected given that the final syllable has a voiceless onset. This now brings us to the second issue investigated in this section. Do voiceless consonants in positions earlier or later than the penult cause stress shift? In (81) it is evident that such stress shift does not occur, as it would create more feet, e.g. (ðσ)(ð)(σ) than the rhythmic (ðσ)(σσ) and hence more alignment violations. Moreover,

<sup>25</sup> I have unfortunately been unable to find an odd-syllable word that would illustrate this pattern (cf. (78c)). However, I tested the proposed grammar with a hypothetical word that would serve as an example and the correct - according to Payne's and Rich's (1988) account - results emerged, i.e. no stress shift.

this is a four-syllable word<sup>26</sup>, hence offering evidence that stress shift only occurs in *odd-syllable* words of the sonorant-voiceless type.

To verify this, let us consider another example with a four-syllable word. Consider the hypothetical example: /kokonaka/. Our system predicts (kòko)(náka), instead of *rightward* stress shift, i.e. (kòko)(nà)(ká) which would produce unwarranted violations of left-edge alignment.

(82) 4σ-word: sonorant - voiceless ---- no rightward stress shift (cf. (72))

ALL-Ft-R = WSP >> ALL-Ft-L

		ALL-Ft-R	WSP	ALL-Ft-L
a.	(kòko)(náka)	** (2)	**	** (2)
b.	(kò)(kòna)(ká)	**** (4)		****! (4)
c.	(kòko)(nà)(ká)	*** (3)	*	*****! (5)

All candidates fare equally well in terms of the high-ranking constraints. The outcome is determined by ALL-Ft-L which chooses the candidate with the fewer feet, since this incurs fewer violations of foot alignment.

But while we have explored the possibility of stress shift to the right as a result of a voiceless consonant occurring in the final syllable, we have not considered the opposite: leftward stress shift (in a manner similar to examples like: [nòwaʃíjáno]) when a voiceless consonant occurs earlier than the penult in the word.

As shown in (83) no such effect arises either. This is exactly a consequence of the equally ranked WSP and ALL-Ft-R. Effectively, stressing syllables with voiceless onsets *early in the word* satisfies the WSP, but simultaneously creates massive violations of right foot alignment. To illustrate the point, some representative candidates with stress shift in positions early in the word are presented below. The same example as in (80) is used.

<sup>26</sup> To be more specific, it is not four-syllable words we are focusing on here, but generally even-syllable words. Two-syllable words would not suffice to illustrate the point, while six-syllable words or more would be inappropriate for expository reasons. This is why I use (hypothetical) examples with four syllables when it comes to even-syllable words.

(83) sonorant - voiceless

ALL-Ft-R = WSP >> ALL-Ft-L

	ALL-Ft-R	WSP	ALL-Ft-L
a. (sàka)(màna)(há)	**** (4)	*	***** (6)
b. (sà)(kà)(màna)(há)	*****!*** (8)		***** (7)
c. (sà)(kàma)(náha)	*****! (6)	*	**** (4)
d. (sà)(kàma)(nà)(há)	*****!* (7)		***** (8)

Candidates (b) and (d) illustrate that although they manage to stress all syllables with voiceless onsets, thus respecting WSP fully, they produce severe violations of right foot-alignment. (c) produces itself a violation of WSP, as the winning candidate does, but it does not fare better than it in terms of ALL-Ft-R. The sum of violations of (a) before it reaches ALL-Ft-L is still smaller than the one of the other contenders so (a) wins.

Summarizing our findings up to this point, it has been shown that the system is generally rhythmic, with ALL-Ft-R quite high-ranked and thus massively violated whenever stress shift occurs. However, in odd-syllable words containing a sequence of voiceless-voiced sonorant in the penult and ultima respectively, high-ranked MORAIC and WSP require that voiceless onsets are not only moraic, but also receive stress. Candidates without stress shift violate WSP to a large extent, while those with stress shift violate ALL-Ft-R more severely. The total violations incurred though are in each case the same, so the decision falls on low-ranked ALL-Ft-L, which favours the stress shifted candidate, as it produces better left alignment (cf. (70)).

In words without the voiceless-voiced sonorant sequence, such stress shift is gratuitous. No trigger exists for stress shift and as a result ALL-Ft-R violations are so many that the optimal candidates present the normal rhythmic pattern. For instance, when the penult-ultima sequence is sonorant-sonorant (cf. (79)), hypothetical stress shift only violates ALL-Ft-R without any gain in terms of WSP simply because voiced sonorant onsets have no moraicity, thus all reasonable contenders - including those that lack stress shift - vacuously satisfy it. Similarly, in a sonorant-voiceless sequence (cf. (80)), the rhythmic pattern ensures that stress will dock on the last syllable, so WSP is satisfied already without any shift. Finally, in voiceless-voiceless sequences (cf. (81)), WSP will be best satisfied by stressing both the penult and final syllables. But if this is done, ALL-Ft-R will be violated to a fatal extent, therefore the system there actually emerges with its normal rhythmic pattern too.



### 2.3.3 Residual issues

#### 2.3.3.1 An alternative: stress shift directly related to primary stress

The reader may have noticed that the stress shift effect only arises in the voiceless-voiced sonorant sequences of *primary stress*. As we have seen already, no similar process occurs in other positions within the word. In the light of this, it is possible to claim that primary and secondary stress are assigned via different algorithms. McGarrity (2003) presents numerous cases where primary stress is assigned differently from secondary stress. The general schema she uses includes a general constraint, e.g. Stress to Weight (STW) and its primary stress variant, i.e. STW<sub>wd</sub>, which makes specific reference to primary stress. Perhaps then, Arabela could be re-analysed along these lines.

A sketch of such an approach is illustrated presently. In accordance with McGarrity, we could invoke the familiar WSP and its primary stress variant WSP<sub>wd</sub><sup>27</sup>.

(84) WSP<sub>wd</sub>: Heavy syllables must have primary stress

The problem though is that in order that WSP (or WSP<sub>wd</sub>) has an effect, the syllable it refers to has to be heavy, because only an unstressed heavy syllable violates it. WSP has no say in unstressed or stressed light syllables. But this means that there must be some constraint which imposes heaviness on syllables. This is MORAIC, the constraint we have already used.

Now, MORAIC does not on its own make any distinction between syllables that would receive secondary or primary stress. It assigns moraicity to as many syllable onsets as it can. Apparently we would like to say that WSP<sub>wd</sub> >> WSP so as to emphasize that primary stress is only affected. The ranking in (55), i.e. \*μ/ONS/[voice] >> MORAIC >> \*μ/ONS, could also be assumed so that only voiceless onsets can be moraic. To simplify things slightly, I will only consider candidates with moraic voiceless onsets, under the view that MORAIC is very highly ranked. Effectively the tableau in (77) now becomes:

---

<sup>27</sup> In fact, McGarrity (2003: 212) uses the constraint PKPROM and PKPROMMAIN for a similar case. She acknowledges that WSP and PKPROM overlap to a certain extent, but uses PKPROM which is satisfied so long as within a word with heavy and light syllables, at least one heavy syllable is stressed. WSP on the other hand requires that every heavy syllable is stressed. This technical detail is not relevant for our purposes, so I am using WSP to be consistent with the previous analysis.

(85) MORAIC >> WSP<sub>wd</sub> >> WSP, ALL-Ft-L

	MORAIC	WSP <sub>wd</sub>	WSP	ALL-Ft-L
a. (s <sup>μ</sup> ap <sup>μ</sup> o)(h <sup>μ</sup> ò)(s <sup>μ</sup> áno)	*	***	*	***** (5)
b. (s <sup>μ</sup> ap <sup>μ</sup> o)(h <sup>μ</sup> òs <sup>μ</sup> a)(nó)	*	****!	**	***** (6)

There are a few points that need discussion here. First note that ALL-Ft-R is not used (although to be accurate, in a word where all syllables are light, it would be needed to generate the right footing)<sup>28</sup>. If it had been used at least in its original position, i.e. next to MORAIC, it would incorrectly produce (b) as the winner. So if it is to be included it has to be at least below the WSP constraints. Now consider WSP<sub>wd</sub>. Both candidates violate it, although the former to a lesser extent, since out of the four syllables with moraic onsets, there is one, i.e. the final, which receives primary stress. On the other hand, candidate (b) fares worse, since none of the moraic onset syllables gets primary stress. Note however, that even in the absence of WSP<sub>wd</sub>, the second candidate would still be excluded. Either the general WSP or ALL-Ft-L would dispense with (b).

This already seems quite suspicious; why should we invoke a constraint that is superfluous? To drive the point home, let us consider one more example. This is the example of a voiced sonorant-voiceless sequence in penult and ultima respectively (cf. (83)). Again let us focus on the more interesting contenders, i.e. the one with stress shift and the other with no stress shift. Both include moraic voiceless onsets.

## (86) sonorant - voiceless

MORAIC >> WSP<sub>wd</sub> >> WSP >> ALL-Ft-L

	MORAIC	WSP <sub>wd</sub>	WSP	ALL-Ft-L
--	--------	-------------------	-----	----------

<sup>28</sup> Marc van Oostendorp (p.c) offers another alternative using WSP<sub>wd</sub>, locating WSP very low in the ranking and without ALL-Ft-R. The proposed ranking is ALIGN-HD<sub>PrWD</sub>-R >> WSP<sub>wd</sub> >> ALL-Ft-L. The first constraint accounts for the fact that rightmost stress is the primary one. This analysis works for examples such as *kòkotáka* and *sàkamànahá* as the interested reader can confirm by drawing the corresponding tableaux. But it will not work for the stress-shifted cases such as: *nòwàfífáno* or *sàpohòsáno* as it predicts instead *nòwàfífáno* and *sàpòhosáno* respectively. I will present one example for illustration purposes. In such cases, the stress shifted candidate has the same number of WSP<sub>wd</sub> violations with the left-aligned one. Since they tie, the decision falls onto ALL-Ft-L which chooses the better left-aligned contender.

(i) Wrong result for stress-shifted cases if the ranking is: ALIGN-HD-R >> WSP<sub>wd</sub> >> ALL-Ft-L

	ALIGN-HD <sub>PrWD</sub> -R	WSP <sub>wd</sub>	ALL-Ft-L
a. (s <sup>μ</sup> ap <sup>μ</sup> o)(h <sup>μ</sup> òs <sup>μ</sup> a)(nó)		****!	***** (6)
b. (s <sup>μ</sup> ap <sup>μ</sup> o)(h <sup>μ</sup> ò)(s <sup>μ</sup> áno)		***	*****! (5)
c. (s <sup>μ</sup> à)(p <sup>μ</sup> òh <sup>μ</sup> o)(s <sup>μ</sup> áno)		***	**** (4)

a.	(s <sup>μ</sup> àk <sup>μ</sup> a)(màna)(h <sup>μ</sup> á)	**	**	*	***** (6)
b.	(s <sup>μ</sup> àk <sup>μ</sup> a)(mà)(náh <sup>μ</sup> a)	**	***!	**	***** (5)

Again WSP<sub>wd</sub> chooses the right winner, but it does not really offer much. The other constraints already provided would still achieve the right winner, although we would need to establish the ranking between WSP and ALL-Ft-L as one of domination of the former over the latter. But all this now looks very similar to the suggested ranking in the previous sections. Although I will not examine the remaining examples of the previous section, it becomes obvious that inclusion of WSP<sub>wd</sub> proves superfluous.

In a final attempt of rescuing this type of analysis, one could make use of general \*μ/ONS/x and the more specific \*μ/ONS/x<sub>wd</sub> or similarly MORAIC and MORAIC<sub>wd</sub>. Both would aim to differentiate between primary and secondary stressed syllables. I will not present analyses that would utilize such constraints here, but it should be evident that even if they could be made to work, they would add unnecessary complexity and duplication to the system, in the sense that a set of different constraints would be needed for primary and for secondary stress. In the absence of supporting evidence for such an approach, it seems to be more economical to derive the effects of this difference indirectly, as it has been proposed in the previous sections.

In other words, it is proposed that stress shift does not relate to the primary-secondary stress distinction, but instead to the combined action of the equally ranked WSP and ALL-Ft-R. The primary-secondary stress distinction is then a side-effect of the above. Stressing syllables with voiceless onsets *early* in the word - this is particularly visible in longer words, e.g. those of 5 syllables - offers a small gain in terms of WSP, but produces massive violations of ALL-Ft-R that cannot be compensated for. Consequently, the rhythmic stress candidate is optimal. In the case of voiceless penult and voiced sonorant ultima however, the language can afford to satisfy WSP more satisfactorily by shifting stress at the expense of only one additional violation of ALL-Ft-R.

### 2.3.3.2 On the peculiarity of the Arabela stress system

The Arabela stress system seems strange because not only it presents effects due to onsets, but perhaps more oddly, this only applies contextually, i.e. when the penult has a voiceless consonant and the ultima has a sonorant. In contrast, in Pirahã all syllables within the trisyllabic window at the right edge which is active for stress, may receive

stress due to their moraicity. Is then this stress shift that Arabela presents as bizarre as one would think? I would like to suggest that it is not.

Supporting evidence comes from languages discussed in Rosenthal and van der Hulst (1999) and Morén (2000), where *codas* are contextually heavy or light. Before elaborating on this, it is worth mentioning that Arabela fits nicely along these cases, the difference being that the contextual factor now affects onsets rather than codas.

What's more, it is largely an analytical issue to suggest that codas (or onsets) are contextually moraic or not (as the above authors do) or that they are consistently moraic, but subject to different constraints making reference to their moraicity, e.g. a distinction between WSP(N) and WSP, where the former is only sensitive to nucleic weight, while the latter to overall weight (as in Pirahã above). Independently of the view taken, the intuition is the same, that is, the moraic behavior of syllables may not be uniform across all the syllables in the word.

As we have seen in detail, in Arabela stress is usually rhythmic and degenerate feet are permitted. Primary stress is thus assigned either on the penult or on the ultima. Stress shift only occurs in odd-syllable words with a voiceless-sonorant penult-ultima sequence. There, instead of the expected final primary stress, this appears on the penult due to the heaviness of that syllable.

A similar case emerges in Goroa (Rosenthal and van der Hulst 1999), this time due to coda weight (87). Stress in Goroa either falls on the leftmost long vowel (a-b) or on the penult (e-f). A further complexity is added by closed syllables which are generally light (b, f) except word-finally (c-d) where they are heavy and can receive stress to satisfy higher ranking metrical constraints, i.e. FTBIN and EDGEMOST-R<sup>29</sup>.

(87)	a.	dú:gnuno:	“thumb”
	b.	girambó:da	“short”
	c.	adúx	“heavy”
	d.	axemís	“hear”
	e.	oromíla	“because”
	f.	idirdána	“sweet”

Stress shift to the right occurs due to the presence of a coda whose heaviness can satisfy EDGEMOST-R. Other languages with this effect, but less reminiscent of Arabela include Eastern Ojibwa, Khalkha Mongolian and Kashmiri (Morén 2000).

---

<sup>29</sup> I am using the constraint as used in the original after Prince and Smolensky (1993/2004). EDGEMOST-L/R requires stress to appear at the left or right edge of the word respectively.

In sum then, the only peculiar thing with Arabela is that stress shift is caused by onsets, and not by codas. Other than that, the emerging pattern is rather ordinary.

#### **2.3.4 Effects on mora affiliation by ranking MORAIC and WSP with accuracy**

A final issue we need to address is the exact output affiliation of moras with segments as regulated by the specific ranking of MORAIC and WSP.

(88)  $*\mu/\text{ONS}/[\text{voice}] \gg \text{MORAIC} \gg * \mu/\text{ONS}$

Through the ranking above, we know that only voiceless onsets are forced to emerge with moras. Things would remain as simple if WSP had no impact on these representations. In a previous ranking (cf. (74)) we had seen that although WSP's exact position was not established, it could at least be positioned next to MORAIC.

A closer look however shows that the exact relationship between the two has further consequences. Importantly, these consequences are merely theoretical and cannot be confirmed empirically one way or another. Both approaches discussed next, generate the right candidate, i.e. the one with stress shift, but differ on whether all voiceless onsets are moraic (cf. (90c)) or only the voiceless onsets of stressed syllables (cf. (89d)).

This difference relates to that fact that to an extent WSP and MORAIC have conflicting demands. MORAIC requires the presence of moras in the output. WSP requires stressing syllables with moraic onsets, but does not penalize syllables with non-moraic onsets if they remain unstressed. This means that it is likely that not all voiceless onsets will surface as moraic, only the stressed ones may do. For this to be possible, it has to be the case that the relevant candidate incurs the same number of violations in terms of the equally ranked MORAIC and WSP with its rival which consists of moraic onsets only. The competition then passes on to  $*\mu/\text{ONS}$ , which naturally chooses the candidate with the fewest moraic onsets. The next example serves as an illustration. This is an example with the by now familiar pattern of voiceless-voiced sonorant onsets in the final two syllables, causing shift of stress to the penult.

(89) MORAIC = WSP = ALL-FT-R >> ALL-FT-L >> \*μ/ONS

/sapohosano/	MORAIC	WSP	ALL-FT-R	ALL-FT-L	*μ/ONS
a. (s <sup>μ</sup> ap <sup>μ</sup> o)(h <sup>μ</sup> os <sup>μ</sup> a)(nó)		** (2)	**** (4)	*****! (6)	**** (4)
b. (s <sup>μ</sup> apo)(h <sup>μ</sup> osa)(nó)	** (2)		**** (4)	*****! (6)	** (2)
c. (s <sup>μ</sup> ap <sup>μ</sup> o)(h <sup>μ</sup> ò)(s <sup>μ</sup> áno)		* (1)	***** (5)	***** (5)	*****! (4)
d. (s <sup>μ</sup> apo)(h <sup>μ</sup> ò)(s <sup>μ</sup> áno)	*(1)		***** (5)	***** (5)	*** (3)

To make the tableau less complex, notice that it can essentially be broken down in a couple of sub-groups<sup>30</sup>. First, candidates (a) and (b) share a rhythmic parsing, whereas (c) and (d) represent the case with the shift of stress. Second, (a) and (c) exhibit moras on all syllables with voiceless onsets, while (b) and (d) only on stressed syllables. The empirical data tell us that either (c) or (d) wins, but we also need to consider the candidates with the normal rhythmic pattern to ensure that the *simultaneous* action of the three constraints [MORAIC, WSP, ALL-FT-R] does not produce any unwelcome surprise by selecting (a) or (b)<sup>31</sup>.

All four candidates incur the same number of violations of the three top-ranked constraints, so lower-ranked constraints will determine the outcome. It is essential that ALL-FT-L dominates \*μ/ONS, so that the first two candidates are excluded. (d) then wins as it incurs one less violation of \*μ/ONS by failing to realize one of the voiceless onset moras. The outcome is the desirable one, but notably it is the candidate, which assigns moras on voiceless onsets only if these receive stress. Unstressed obstruent onsets have no moras.

As mentioned, the ranking ALL-FT-L >> \*μ/ONS is also established. If it were \*μ/ONS >> ALL-FT-L, then (b) would be wrongly selected. If the ranking was undetermined, i.e. ALL-FT-L, \*μ/ONS, no single winner would be chosen, since both (b) and (d) would produce eight violations of the relevant constraints in total, versus ten of (a) and nine of (c).

The alternative would be to let our grammar choose the output where all voiceless onsets are moraic irrespective of whether they are onsets of stressed or unstressed syllables, i.e. as in (89c). Simply ranking MORAIC over WSP achieves that.

<sup>30</sup> Another plausible candidate is (s<sup>μ</sup>à)(p<sup>μ</sup>ò)(h<sup>μ</sup>ò)(s<sup>μ</sup>áno) which would perfectly satisfy MORAIC and WSP. Nonetheless, it would incur nine violations of ALL-FT-R which would prevent any further evaluation.

<sup>31</sup> It is again important that these three constraints act simultaneously and are equally ranked. Alternatively, given that WSP = ALL-FT-R, we would have MORAIC, WSP = ALL-FT-R, which would entail either MORAIC >> WSP = ALL-FT-R or WSP = ALL-FT-R >> MORAIC. The former would produce (89c) as the winner, which is a reasonable possibility (cf. (90c)), but the ranking WSP = ALL-FT-R >> MORAIC would give us (89b) which is blatantly wrong. Therefore, equal ranking is once more crucial.

We will see that given this ranking, the precise ranking of  $*\mu/\text{ONS}$  and ALL-Ft-L becomes irrelevant.

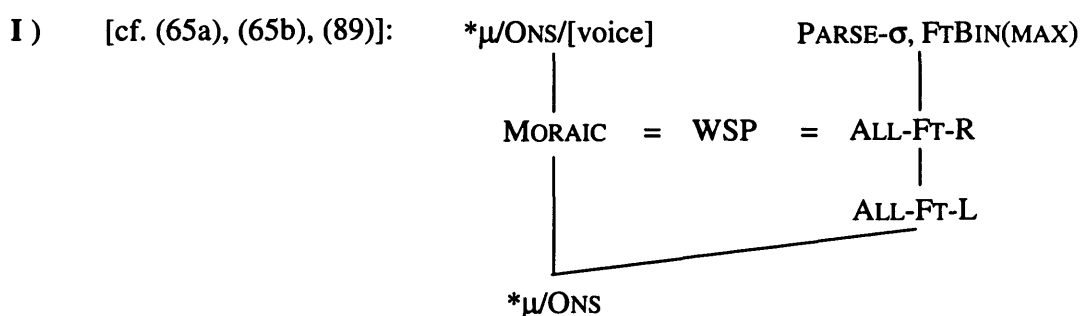
(90) MORAIC >> WSP = ALL-Ft-R >> ALL-Ft-L,  $*\mu/\text{ONS}$

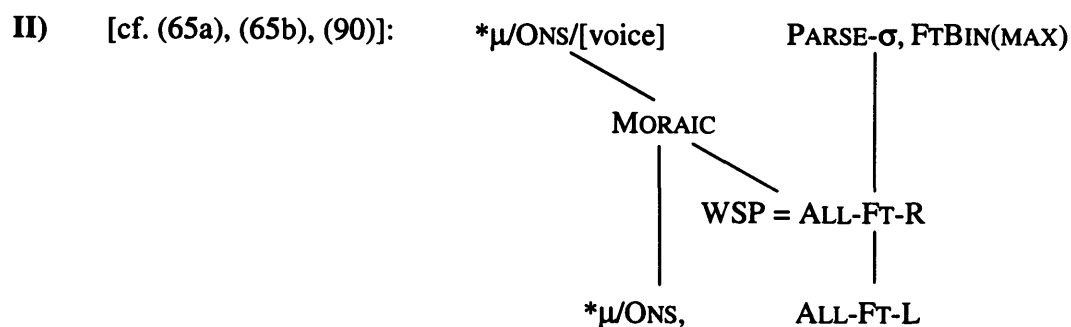
/sapohosano/	MORAIC	WSP	ALL-Ft-R	ALL-Ft-L	$*\mu/\text{ONS}$
a. (s <sup>μ</sup> ap <sup>μ</sup> o)(h <sup>μ</sup> òs <sup>μ</sup> a)(nó)		** (2)	**** (4)	***** (6)	***!(4)
b. (s <sup>μ</sup> apo)(h <sup>μ</sup> òsa)(nó)	**! (2)		**** (4)	***** (6)	** (2)
c. (s <sup>μ</sup> ap <sup>μ</sup> o)(h <sup>μ</sup> ò)(s <sup>μ</sup> áno)		* (1)	***** (5)	***** (5)	**** (4)
d. (s <sup>μ</sup> apo)(h <sup>μ</sup> ò)(s <sup>μ</sup> áno)	*! (1)		***** (5)	***** (5)	*** (3)

Here (b) and (d) - our previous winning candidate - lose early on because they fail to assign moras on all obstruent onsets. The remaining candidates tie on the next two constraints, so the decision is passed on the very low-ranked ones. The winner (c) is finally chosen. As in (89), this is the one that exhibits shift of stress. Nonetheless, it keeps all moras on the voiceless onsets, in contrast to (89d), which retains only the voiceless onset moras of stressed syllables.

At this stage, there is no principled reason that should force us choose one over the other grammar. Perhaps, a simpler system is produced through (90) in the sense that all voiceless onsets are uniformly treated as moraic and that might be a property more easily acquired by the language learner, rather than being based on stress cues to decide which voiceless onset moras survive in the output. On the other hand, the system in (89) is consistent with other approaches mentioned in the previous section (cf. Morén 2000, Rosenthal and van der Hulst 1999) which suggest that certain (coda) consonants are variably moraic; for instance according to Morén (2000), codas in Kashmiri are moraic when they are part of stressed syllables, but emerge as non-moraic in unstressed syllables. With this consideration in mind, we can conclude that the system proposed for Arabela is one of the following:

(91) Mini-grammar of Arabela - Final attempt





### 2.3.5 Arabela Conclusion

In this section data from Arabela, a very little-studied Peruvian language, have been presented, and I have attempted to give a detailed account of the language's stress system. Stress is largely rhythmic forming trochees (binary or degenerate) from left to right. In the event of odd-syllable words with voiced/sonorant onsets in the ultima and voiceless ones in the penult, stress shifts to the penult instead of its anticipated assignment on the final syllable. An account has been proposed where this is a result of the interaction between the language's normal stress system and constraints that refer to onset moraicity.

The disruption of the normal rhythmic stress algorithm occurs when the penult has a voiceless and the ultima a voiced sonorant onset, because only there can the foot-alignment constraints and WSP be satisfied in the best possible way by shifting the stress to the penult. In all other instances, such a shift would be gratuitous causing unnecessary violations of ALL-Ft-R.

## 2.4 Karo

Karo is the third language I will discuss in this chapter. It is a Tupi language spoken in the Rondônia State, Brazil by approximately 150 Arara Indians. The data presented here are those that Nilson Gabas Jr. has collected during field work and published in Gabas (1998) and Gabas (1999), henceforth referred to as G98<sup>32</sup> and G99 respectively. Here too, stress is sensitive to the quality of the onset. In fact, Karo suggests a system where sonorants and voiceless obstruents pattern as moraic onsets to the exclusion of voiced obstruents which are non-moraic. This is sharply different from what we have seen in Pirahã and Arabela, but still explainable in the current system. Stress, as we will also

<sup>32</sup> G98 is written in Portuguese. As I have not attempted any translation of the glosses into English, these are given in the original language.



see, is sensitive to tone and nasalization as well. But first, let us present the segmental inventory of the language. Throughout, H tone is marked with an acute accent, L tone has no accent at all, the tilde underneath the vowel represents nasality and finally boldface represents stress.

### 2.4.1 Karo segmental inventory

I begin by presenting the phonemic consonantal inventory first.

#### (92) *Karo phonemic consonants*

	<b>bilabial</b>	<b>alveolar</b>	<b>palatal</b>	<b>velar</b>	<b>glottal</b>
<b>Stops</b>	p b	t r	c	k g	ʔ
<b>Nasal</b>	m	n		ŋ	
<b>Fricative</b>					h
<b>Approximant</b>	w		y		

Two points are worthy of discussion here. First, the language treats /r/ (which is phonetically a flap [ɾ]) as a /d/ phonologically. Gabas shows that [ɾ] behaves phonologically and morphologically as the voiced stops [b] and [g] do and thus is included among them. Alternations across words like the one in (93) provide evidence for this fact.

#### (93) *Data from G98: 22*<sup>33</sup>

pãt + ʔaʔ → pã:r aʔ (/t/ turns into [ɾ])  
 up + ʔaʔ → u:b ʔaʔ (/p/ turns into [b])<sup>34</sup>  
 tak + ʔip → ta:g ip (/k/ turns into [g])

This property is not unprecedented; other Tupi languages treat /b r g/ as the voiced counterparts of /p t k/ (Rodrigues p.c. to Gabas 1999: 12), while in Gadsup, Papua New Guinea (Frantz and Frantz 1966: 3), underlying voiced alveolar *d*, surfaces as either *d* or *r* intervocalically. I will thus share this idea too.

The second point is that the only fricative in the language is /h/, which also is less frequent than other phonemes (G99: 12). Phonetically, the following consonants surface.

<sup>33</sup> Vowel length is not distinctive in Karo. This lengthening here is optional. See Gabas (1998: 22) for details.

<sup>34</sup> Unlike the other examples, the initial glottal of /ʔaʔ/ here remains. Gabas does not discuss why this is the case.

(94) *Karo phonetic consonants*

	bilabial	alveolar	palatal	velar	glottal
<b>Stops</b>	p p: p' b	t t: t'	c c:	k k: k' g	ʔ
<b>Flap</b>		r r̃			
<b>Nasal</b>	m m <sup>b</sup> <sup>b</sup> m	n n <sup>d</sup> <sup>d</sup> n		ŋ ŋ <sup>g</sup> <sup>g</sup> ŋ	
<b>Fricative</b>	β		ç	ɣ	h
<b>Approximant</b>	w w̃		y ÿ		

Some remarks that will also facilitate the presentation of the data below are in order. Long voiceless stops appear as onsets of stressed syllables. Unreleased voiceless stops occur as final codas. In all other contexts, we get the plain alternants. Nasal sonorants appear as pre-stopped nasals, e.g. <sup>b</sup>m, in word-final codas of stressed syllables with an oral vowel, whereas as post-stopped nasals, e.g. m<sup>b</sup>, in onsets of stressed syllables that include an oral vowel. Plain nasals occur everywhere else. [b]~[β] and [g]~[ɣ] are in variation as onsets of unstressed syllables preceded by an open syllable, while [c]~[ç] are in entirely free variation. Finally, the nasalization of the approximants and the flap is predictable based on the nasalization of surrounding vowels.

The phonemic vowels of Karo include: i) the oral i, e, ɪ, ə, a, u and o and ii) the nasal j, ɛ, ɶ, and ɔ. Phonetically, two points are important. The first is that the oral mid vowel /e/ surfaces as: i) [e] in H-toned syllables, ii) [ɛ] in unstressed or L-tone stressed syllables. Somewhat similarly, /o/ surfaces as: i) [o] in H-toned syllables and in unstressed syllables, but as ii) [ɔ] in L-tone stressed syllables. The second point is that the nasalized /ə/ surfaces as [ɶ].

## 2.4.2 Karo tone, nasalization and stress

With this much segmental information at hand, we may proceed in looking at Karo in more detail. Syllables are of the (C)V(C) type. No long vowels seem to arise, whereas sequences of vowels are syllabified in different syllables. Most words do not exceed three syllables in length.

Karo distinguishes between H and L tone<sup>35</sup>. H tone is marked by the acute accent, and L tone by the absence of any mark. Apart from having phonemically nasal segments, the language also possesses rules that assign nasalization to neighbouring segments (see G98: 63).

<sup>35</sup> Minimal pairs include for example: [pɛ́n] 'to open' vs. [pɛn] 'to step', [cá<sup>d</sup>n] 'to wash' vs. [ca<sup>d</sup>n] 'to pluck' (G99: 44).

Stress is generally final, but can be pushed back to an earlier position if one of the following conditions is applicable in the order specified here: i) if a syllable has H tone, then it gets stress, ii) if there is no H-toned syllable, but there is one with a nasal vowel, then this receives stress and finally iii) if none of the above is applicable, but if the final syllable has a voiced stop onset then the penult gets stressed<sup>36</sup>. The overwhelming majority of words present stress in a disyllabic window at the right edge. Only rarely, antepenult stress is found, but the conditions where this occurs are not clear (cf. fn. 38 for discussion).

Although my focus will be on the final condition, I will briefly discuss the other two for completeness<sup>37</sup>. The relationship between tone and stress has been explored in de Lacy (2002), who shows that H-toned syllables commonly attract stress. As far as nasalization is concerned, John Hajek (p.c.) suggests that in fact nasalized vowels may be simply long, in which case their primacy for stress falls out easily (although some reference to WSP(N)) would be needed to assign more importance to weight coming from nasalized vowels, compared to other moraic segments).

As implied above, stressing a syllable with H-tone is top priority, even at the expense of stressing a final syllable with a voiced stop onset (95a, b).

(95) *Stress and tone*

- |    |                                     |                |           |
|----|-------------------------------------|----------------|-----------|
| a. | yogá                                | ‘egg’          | (G99: 43) |
| b. | korét’                              | ‘guan (sp.)’   | (G99: 43) |
| c. | man <sup>d</sup> ogo <sup>d</sup> n | ‘rabbit (sp.)’ | (G99: 40) |

In the absence of a H-toned syllable, nasalization attracts stress, again even at the expense of stressing a final syllable with a voiced stop onset (96d, e)<sup>38</sup>.

<sup>36</sup> I will discuss the case of a voiced stop onset in the penult too at a later point.

<sup>37</sup> H-tone and nasalization spreading occur in Karo under certain circumstances. Certain data seem to suggest that stress assignment precedes both. For instance, the underlying form for [nayúá] is /nayúá/ ‘ant’s house’ and for [wífup] it is /wífup/ ‘native, non-domesticated’ (G99: 43). Since default stress is final, we would expect stress on the final instead, given that this also has H tone. This is not the case though. The same point obtains for nasalized vowels too. In addition, an optional rule turns oral vowels to their nasalized counterparts when they are between nasals (G98: 63), e.g. /mani/ → [mani] ~ [māni] ‘macaxeira’; /anana/ → [anan<sup>d</sup>a] ~ [anān<sup>d</sup>a] ‘abacaxi’. Again, stress seems to precede nasalization otherwise we should get \*[māni] or \*[anān<sup>d</sup>a].

<sup>38</sup> Some other data need to be treated with caution. Gabas seems to suggest that there is a disyllabic window at the right edge of the word, which may receive stress. I have however found three examples where this is violated by allowing antepenultimate stress on iyáno?n<sup>d</sup>on<sup>d</sup> ‘machado’ (G98: 16), /agóá?pət/ → [agóá?pət] ‘pajé’ and /cǎbe?/ → [cǎbe?] ‘bow’ (G98: 73). One could perhaps attribute this to the fact that stressing an input H-tone is of such priority that it can override the window. Some other examples that exhibit nasalization seem problematic in the light of Gabas’ claim (G99: 42) that there can only be just one underlying nasal per word, and yet here we have two in words such as: /papǎ/ → pǎpǎ ‘yam (sp.)’ [G99: 23], /mǎygara/ → mǎygǎra ‘snake’ [G99: 17]. Furthermore G99: 43 argues

(96) *Stress and nasalization*

a.	maʔo	‘ant (sp.)’	(G99: 42)
b.	pəŋa	‘dar’	(G98: 17)
c.	moŋja	‘miçanga’	(G98: 39)
d.	piŋon	‘redondo’	(G98: 30)
e.	carekʔ	‘slow’	(G99: 23)

### 2.4.3 Stress and onset voicing

#### 2.4.3.1 The core analysis

Setting tone and nasalization aside, we can focus on the cases where none is present in the word, thus stress is normally word-final, unless that syllable has a voiced obstruent onset in which case stress retracts to the penult. Both patterns are illustrated below.

(97) *Karo stress and onset voicing* (G99: 39-41)

a.	maʔpe	‘gourd’	(G99: 14)
	kəyɔ	‘crab’	
	yaʔm <sup>b</sup> ɔ	‘yam (sp.)’	
	pakɔ	‘fontanel’	
b.	kɪrɪwɛpʔ	‘butterfly’	
	kuruʔcu	‘saliva’	
c.	yaba	‘rodent (sp.)’	
	pibɛʔ	‘foot’	
	wɛɛ	‘frog’	
	karo	‘macaw’	
	məga	‘mouse’	
	iʔcəɔ	‘quati (sp.)’	

(97a) presents cases with final stress. The examples here involve all combinations of sonorants and voiceless obstruents in the two final syllables of the word. Stress is consistently final. This is also the case in (97b). This time, the penult includes a voiced stop onset, whereas the final has either a sonorant or a voiceless stop. Things are different in (97c). Here the *final* has a voiced stop, and the penult has either a voiceless stop or a sonorant. Stress is now systematically penultimate.

---

that “high pitch never occurs in a syllable of a word which also contains another syllable with an underlying nasal vowel”. Several counterexamples arise though:

i)	/cəabeʔ/	cəabeʔ	‘bow’ (G98: 73)	/páram/	páɾɾám	‘wood (sp.)’ (G99: 46)
	/káram/	káɾɾám	‘hummingbird’ (G99: 46)	/páŋan/	páŋɛn	‘to give’ (G99: 46)
	/kéŋan/	kɛɾɛn	‘dormiram’ (G98: 80)		érɔm	‘lamber’ (G98: 24)

To analyse this set of data, I will again make reference to onset moraicity. Recall that in Pirahã and Arabela, we had claimed that only voiceless obstruents can be moraic when in onset position. Voiced obstruents and sonorants had no moraicity. For Karo, the proposal is that voiceless obstruents *and sonorants* are moraic onsets, whereas voiced obstruents are not.

Although this might seem a striking conclusion at first, it is not really, if we recall the discussion about the relation between pitch and sonorants that appeared in chapter §1.4.1.1. There, I offered tonal cases where sonorants could either behave like the voiced obstruents or like the voiceless counterparts<sup>39</sup>. This dual nature of sonorants was attributed to the fact that sonorants are inherently voiced and by default lack the [voice] feature characteristic of voiced obstruents. This is not to say that sonorants cannot bear the [voice] feature at all. In fact, they can, which is why in some languages they pattern with voiced obstruents. But when they lack the [voice] feature, then it is quite likely that they behave in a similar way to voiceless obstruents<sup>40</sup>.

The idea was then to extend the effect of pitch raising due to (lack of) voicing beyond tone and propose that some languages construe this as stress. As with tone, in stress too, sonorants should be expected to demonstrate their dual status. In Pirahã and in Arabela we saw them patterning alongside voiced obstruents. Now in Karo, we see them behaving like voiceless obstruents do. Karo sonorants thus lack a [voice] feature, which is why in the ranking in (98), \*μ/ONS/[voice] does not refer to them, allowing sonorants to be moraic as voiceless obstruents are.

(98) \*μ/ONS/[voice] >> MORAIC >> \*μ/ONS

---

<sup>39</sup> G98: 80-84 claims that sonorants and voiced obstruents pattern together as they allow H-tone spreading across them, whereas voiceless obstruents block such spreading. This finding is unsatisfactory for two reasons: first, as Gabas himself acknowledges, it is cross-linguistically an unlikely pattern since one would expect voiced obstruents blocking the H-tone spreading given that these are standardly depressors. But even if this were possible, a more serious problem is that the examples with voiceless obstruents Gabas gives are really not comparable with their sonorant and voiced obstruent counterparts. For voiced obstruents and sonorants, stress is penultimate and there is an input H-tone on the same syllable. In the output, the H-tone spreads onto the final, e.g. /o/+/pitégat/ → [owit:égát'] and /o-/+/káwan/ → [ok:áwán]. On the other hand, the examples with voiceless obstruents involve final stress and no lexical H tone whatsoever making spreading inherently impossible, e.g. /caʔpe/ → [caʔpe] or /moca/ → [mocia]. I believe it would be hazardous to draw any conclusions from such unclear data.

<sup>40</sup> Karo sonorants are not simply inert. They actively attract stress away from syllables with voiced obstruents, as pointed out through a word like: mæga 'mouse'. If they were inert, then we would only see stress retraction when the final syllable had a voiced onset and the penult had a voiceless one, but not when the penult had a sonorant one.

Thus, while stress is generally assigned to the final syllable (99i), whenever this syllable includes a voiced onset and the penult includes either a voiceless obstruent or sonorant onset, stress shifts to the penult (99ii).

(99) *Karo stress*<sup>41</sup>

- i. General pattern:  $\sigma$   $\sigma_{[\neq \text{voi}]}$
- ii. Stress shift:  $\sigma_{[\neq \text{voi}]}$   $\sigma_{[= \text{voi}]}$

The way this can be accounted for is that stress is restricted within a disyllabic window at the right edge, in which WSP is active and dominates the preference for final stress captured by ALIGN-HD-R (also used previously in Pirahã).

(100) WSP >> ALIGN-HD-R

WSP is violated if syllables with moraic onsets receive no stress. In particular, when the ultima has a voiced non-moraic onset and the penult a non-voiced moraic one, then penult stress appears as a better candidate, since WSP is satisfied. This point is schematically presented below (101).

(101) *Penult stress as a result of stress shift* [syllables with moraic onsets have an indexed mora]

	$\sigma_{[\neq \text{voi}]}$	$\sigma_{[= \text{voi}]}$	WSP	ALIGN-HD-R
a.	$\sigma_{\mu[\neq \text{voi}]}$	$\sigma_{[= \text{voi}]}$		*
b.	$\sigma_{\mu[\neq \text{voi}]}$	$\sigma_{[= \text{voi}]}$	*!	

When the ultima has any other type of onset, such stress shift does not occur, as there is no trigger for its application. In particular, if the penult has a voiced onset [cf. (102)], then both WSP and ALIGN-HD-R are satisfied more satisfactorily in final stress (102a) than in penult stress (102b).


(102) *Final stress as a result of simultaneous WSP and ALIGN-HD-R satisfaction*

	$\sigma_{[= \text{voi}]}$	$\sigma_{[\neq \text{voi}]}$	WSP	ALIGN-HD-R
a.	$\sigma_{[= \text{voi}]}$	$\sigma_{\mu[\neq \text{voi}]}$		
b.	$\sigma_{[= \text{voi}]}$	$\sigma_{\mu[\neq \text{voi}]}$	*!	*

<sup>41</sup> To simplify representations, I use [= voi] to mark voiced obstruents and [ $\neq$  voi] to collectively refer to sonorants and voiceless obstruents.

In the other cases which involve all the combinations of voiceless obstruents and sonorants in the final two onsets, WSP is equally satisfied by both candidates. Consequently, ALIGN-HD-R picks out the candidate with final default stress [cf. (103)].

(103) *Final stress as a result of equal WSP satisfaction. ALIGN-HD-R decides*

	$\sigma_{[\neq \text{voi}]}$	$\sigma_{[\neq \text{voi}]}$	WSP	ALIGN-HD-R
 a.	$\sigma_{\mu [\neq \text{voi}]}$	$\sigma_{\mu [\neq \text{voi}]}$	*	
b.	$\sigma_{\mu [\neq \text{voi}]}$	$\sigma_{\mu [\neq \text{voi}]}$	*	*!


This ranking also makes a further prediction. Recall in the beginning of this section that penultimate stress occurs when the final onset is a voiced obstruent. This is actually the claim that G99: 40 makes. It thus implies that penult stress should occur even when the penult has itself a voiced obstruent onset. Gabas nonetheless provides no example of this sort. What is more he presents two examples which he treats as exceptions to this generalization. These are:

(104)  $\sigma_{[= \text{voi}]} \sigma_{[= \text{voi}]}$  words: *exceptions for Gabas, norm for current account* (G99: 41)

a.	kiribɔp'	*kiribɔp'	'frog (sp.)'
b.	miririy	*miririy	'toad (sp.)'

Notably, both penult and ultima have voiced onsets. And yet, final stress arises. While this is an exception to Gabas' statement, it is in fact anticipated given the analysis sketched here. The reason stress shift does not occur is because both candidates vacuously satisfy WSP, therefore ALIGN-HD-R is again the determining factor in favour of final stress (105a).

(105) *Voiced onsets: Final stress because ALIGN-HD-R decides*

	$\sigma_{[= \text{voi}]}$	$\sigma_{[= \text{voi}]}$	WSP	ALIGN-HD-R
 a.	$\sigma_{[= \text{voi}]}$	$\sigma_{[= \text{voi}]}$		
b.	$\sigma_{[= \text{voi}]}$	$\sigma_{[= \text{voi}]}$		*!

In sum then, the part of the Karo stress algorithm that bears on the onset quality issue reveals that sonorants pattern alongside voiceless obstruents in being moraic onsets. The ranking of  $*\mu/\text{ONS}/[\text{voice}] \gg \text{MORAIC} \gg *\mu/\text{ONS}$  (cf. (98)) combined with the assumption that sonorants lack the [voice] feature produces this result. The second major part of the analysis involves the ranking  $\text{WSP} \gg \text{ALIGN-HD-R}$  (cf. (100)), which requires that heavy syllables get stress. When the ultima has a voiced onset and the

penult another type of onset, then the penultimate syllable receives stress as it is heavy. When the ultima has any other type of onset, then final stress simultaneously satisfies WSP (by stressing a heavy syllable) and right-edge alignment of stress. Similarly, if both ultima and penult have voiced onsets, stress is again word-final by vacuous satisfaction of WSP and perfect stress alignment.

### 2.4.3.2 Remaining issues

#### 2.4.3.2.1 Onsetless syllables

As mentioned in the beginning, Karo allows onsetless syllables, both word-initially and word-medially.

#### (106) *Onsetless syllables in Karo*

a.	i.yə	‘Brazil nut’	(G99: 18)
	i.ci	‘water’	(G99: 19)
	a.me.kɪ	‘jaguar’	(G99: 21)
b.	wí.up	‘native, non-domesticated’	(G99: 43)
	có.a.be?	‘bow’	(G98: 73/G99: 25)
c.	é.rɔm	‘lamber’	(G98: 24)
	j.ya	‘bird’	(G99: 42)
d.	i.ɔt	‘pescoço’	(G98: 28)
	e.i	‘irara’	(G98: 22)

(106a) exemplifies onsetless syllables word-initially, and (106b) word-medially. Stress is assigned as previously, so that syllables with H-tone or nasalization attract stress. The point made in (106c) is that onsetless syllables may receive stress if they are the best possible stress bearers given that they also carry tone or nasalization. More importantly, (106d) shows that even in the absence of these, onsetless syllables can still carry stress, without any need to insert an onset.

As a result, onsetless syllables in Karo do not push stress away verifying a previous claim that appeared in the previous chapter and repeated in the very beginning of the present one (cf. (1)), namely that the effects from the presence of the onset on stress are independent from those of the quality of the onset on stress. Karo manifests the case where the quality of the onset matters - as argued for in the previous section - but where the presence of the onset is unimportant. Onsetless syllables may carry stress in the same way onsetful ones do.



#### 2.4.3.2.2 Exceptions to the stress shift pattern and the role of codas

Gabas (1999: 41) lists ten examples which present themselves as exceptions to the stress shift pattern, as they fail to undergo shift, although we would expect them to do so.

The first two examples have already been discussed. They were the cases where both the penult and the ultima contain a voiced obstruent, but stress does not shift. The examples are repeated from (104).

(107) *Exceptions for Gabas, norm for current analysis:  $\sigma_{[= \text{voi}]}$   $\sigma_{[= \text{voi}]}$  words:*

- |    |                 |                  |              |
|----|-----------------|------------------|--------------|
| a. | <b>kiribəp'</b> | <b>*kiribəp'</b> | 'frog (sp.)' |
| b. | <b>miririy</b>  | <b>*miririy</b>  | 'toad (sp.)' |

As explained before, this pattern is exactly what we expect in the current analysis, therefore these are no longer exceptions. The next three examples are indeed genuine exceptions given what we have said so far. We may either accept that this is the case or maybe speculate that perhaps final syllables here possess H tone or nasalization that has not been recognized. After all, certain data, as mentioned in fn. 38, are not as clear or present some inconsistencies, so this is not an unreasonable possibility.

(108) *Truly exceptional cases (G99: 41)<sup>42</sup>*

/acibe/	[acibe]	<b>*[acibe]</b>	'root'
/pobo/	[pəbə]	<b>*[pəbə]</b>	'owl'
/cagəp/	[cagəp']	<b>*[cagəp']</b>	'dish'

The final five examples are below<sup>43</sup>. We will deal with these more extensively.

(109) *Exceptions that involve final sonorant codas (G99: 41)*

/korem/	[kore <sup>b</sup> m]	<b>*[kəre<sup>b</sup>m]</b>	'also'
/kora/	[kora <sup>d</sup> n]	<b>*[kəra<sup>d</sup>n]</b>	'fish (sp.)'
/pirun/	[piru <sup>d</sup> n]	<b>*[pɪru<sup>d</sup>n]</b>	'round'
/pagon/	[pagə <sup>d</sup> n]	<b>*[pəgə<sup>d</sup>n]</b>	'friend'
/yogoy/	[yogəy]	<b>*[yəgəy]</b>	'breath'

The common factor here is that the final coda comprises a sonorant consonant. It is thus possible to propose that these codas are moraic. If they are, then the moraic composition

<sup>42</sup> Two more examples are exceptional given the current proposal: [aəɔ] instead of [**\*aəɔ**] 'parrot' (G99: 39) and [yaɪ] rather than [**\*yaɪ**] 'macaco guariba' (G98: 25).

<sup>43</sup> Thanks to Moira Yip (p.c.) for suggesting closer examination of these data.

of a word like [yogɔy] is the following [y<sup>u</sup>o<sup>u</sup>.gɔ<sup>u</sup>y<sup>u</sup>]. Since both syllables are bimoraic, they are equally eligible for stress. ALIGN-HD-R can now play a role in selecting final stress.

To see whether this analysis is viable, we need to consider the issue of codas in more detail. Word-medially, one only finds the coda /ʔ/ and much less frequently /h/ (G99: 12, fn. 6), e.g. ihɣɔy ‘piranha’ (G99: 18)<sup>44</sup>. The fact that final open syllables receive stress although they are being preceded by closed ones, e.g. maʔpe, should make it obvious that these codas could not be moraic. Had they been moraic, then in a word like [maʔpe], every segment would carry a mora, making the first syllable trimoraic and the final one bimoraic. Due to the ranking WSP >> ALIGN-HD-R, and the gradient computation of WSP (cf. the Pirahã analysis), we would thus expect penult stress.

(110) *Non-moraic codas word-medially*

maʔpe	‘gourd’	(G99: 14)
naʔtɔ	‘tapir’	(G99: 14)
piʔti	‘heavy’	(G99: 18)

Word-finally, one finds a larger set of codas. They can be unreleased stops, the glottal stop (primarily as a coda of monosyllabic functional words) or sonorants (discussed below).

(111) *Final codas: non-sonorants*

pewítʔ	‘honey, sweet’	(G99: 13)
naʔcəkʔ	‘hole’	(G99: 14)
makɪapʔ	‘peanut’	(G99: 17)
ʔaʔ	CLASSIFIER	(G98: 13)

The non-moraicity of non-sonorants (cf. (110)) can also be seen word-finally, as exemplified below.

<sup>44</sup> I found two examples with a sonorant coda word medially; both included the sequence /mɔy/ as in mɔyɣɔɣɔ (G99: 42) and mɔyhi (G98: 38). Observe that both have unconventional stress. In the first case, one might expect stress on the antepenult (in analogy to [agóáʔpətʔ]) and in the second, stress on the penult given that it is nasalized. To understand why antepenult stress could be anticipated, recall from footnote 38 that there are a handful of cases where stress is antepenult if in that position there is a H-tone or a nasalized vowel. Moreover, it seems to be the case that stress is assigned before nasalization or H-tone spreading applies. If this is indeed the case, then we should assume that nasalization in mɔyɣɔɣɔ originates in the first syllable (and then spreads rightward; there is no leftward spreading), and since stress assignment precedes spreading, we would expect antepenult stress.

(112) *Non-moraic non-sonorants*

cɪɾɪpʔ	‘bird (sp.)’	(G99: 39)
pibeʔ	‘foot’	(G99: 39)

The claim is that default stress is final. Voiceless obstruent onsets contribute a mora to their syllable, thus the first syllable of these words is bimoraic. Had the non-sonorant coda added a mora, then the second syllable should be bimoraic too. Thus, if they were equal in weight, we would expect final stress due to better stress alignment. Penultimate stress shows that the final coda adds no mora.

On the other hand, whenever a sonorant is a final coda, stress is word final, unless the penult has a H tone or a nasal V, which is in accordance with the fact that weight submits to both.

(113) *Priority of H-tone or nasalization over weight*

/yogqyom/	yogqyom	‘beard, moustache’	(G99: 42)
/kɔnam/	kɔnəm	‘crazy’	(G99: 42)
/káwan/	káwán	‘be fat’	(G99: 46)

But whenever H-tone or nasalization are not present and in contrast to final non-sonorant codas, we can see that final sonorant codas attract stress by virtue of being bimoraic. Final stress will thus be preferred there because of better stress alignment. This is supported by the data that Gabas presents as exceptions to his generalizations, repeated as (114).

(114) *Final sonorant codas are moraic*

/korem/	[kore <sup>b</sup> m]	*[kəre <sup>b</sup> m]	‘also’
/kora/	[kora <sup>d</sup> n]	*[kəra <sup>d</sup> n]	‘fish (sp.)’
/pirun/	[piru <sup>d</sup> n]	*[pɪru <sup>d</sup> n]	‘round’
/pagon/	[pagɔ <sup>d</sup> n]	*[pəɔ <sup>d</sup> n]	‘friend’
/yogoy/	[yogɔy]	*[yɔgɔy]	‘breath’

The proposal then is that codas in Karo are moraic as predicted by Zec’s (1988, 1995) system. Sonorant codas are moraic, whereas non-sonorant ones are not (cf. Kwakwala). Karo is then unique in that it simultaneously presents the relationship between well-formedness, sonority and moraicity in both codas and onsets. The pattern in (114) can

easily be accounted for with the existing constraints and under the assumption that sonorous codas are moraic<sup>45</sup>.

(115) *moraic sonorant codas: WSP >> ALIGN-HD-R*

	yogɔy	WSP	ALIGN-HD-R
a.	y <sup>μ</sup> o <sup>μ</sup> gɔ <sup>μ</sup> y <sup>μ</sup>	*	
b.	y <sup>μ</sup> o <sup>μ</sup> gɔ <sup>μ</sup> y <sup>μ</sup>	*	*!

This has an additional welcome result. Without the proposal about the partial moraicity of codas, we would have no explanation why the following two words receive final stress.

- (116) pɛɔ<sup>d</sup>n            \*pɛɔ<sup>d</sup>n            ‘skin’            (G99: 40)  
           penaɔy            \*penaɔy            ‘dançar’        (G98: 16)

Without final moraic sonorant codas, we would expect final stress, because the final syllable would have just one mora. The penult would have two due to the onset and nucleus moras, therefore stress should be dragged on to that syllable. The fact that it does not is difficult to account, but not if sonorant codas contribute moras. In that case, the final syllable is bimoraic too, consequently stress remains put as alignment is perfect.

#### 2.4.4 An alternative: stress is lexical, voicing is predictable (Blumenfeld 2005b)

The exploration of Karo stress has been recently the object of study in Blumenfeld (2005b). To my knowledge, his analysis and the present one are the only detailed theoretical accounts for this set of data. Blumenfeld puts forward an analysis which argues that the stress system of Karo is lexically determined and the voicing of segments is instead predictable, contra Gabas (1998, 1999) and the current analysis, where stress is predictable and voicing is contrastive. In a nutshell, Blumenfeld’s account builds on the fact that voiced and voiceless stops only contrast in stressed syllables. But, although Blumenfeld claims that he does away with the voicing contrast,

<sup>45</sup> I have been able to find one exception to this pattern, namely the word: [ɛci:igu<sup>d</sup>n] ‘ant-eater (sp.)’ (G99: 40) instead of the anticipated \*ɛci:igu<sup>d</sup>n.

in fact he does not, since this is still needed in stressed syllables. This proves a weak point compared to the present proposal, where indeed only one contrast exists, that of voicing, leaving stress to be regulated by certain principles.

Blumenfeld furthermore argues that in all positions other than the stressed ones, voiceless stops emerge, except in intervocalic onsets of stressless syllables. There, the voiced counterparts arise instead as a product of intervocalic lenition. Other morphophonemic processes in which voiceless stops become voiced provide support for lenition, e.g. as in the context [V]\_[V]<sub>[-accent]</sub> where the vowels belong to different morphemes, e.g. /e-tati/ → [erati] ‘te trouxe’.

(117) *Karo voicing facts as summarized in Blumenfeld:*

- Voicing only contrasts in stressed syllables, e.g. *mat:et* ‘yesterday’ vs. *korét* ‘bird (sp.)’ [G99: 17]
- Elsewhere, voiceless stops occur except in intervocalic onsets of stressless σ
  - a. V\_’V (in this position voiceless stops are lengthened), e.g. *mat:et*
  - b. ?\_V e.g. *pi?ti* \**pi?ri* ‘heavy’ [G99: 18]
  - c. #\_V e.g. *pewit* \**bewit* ‘honey’ [G99: 45]
  - d. V\_# e.g. *na?cək* \**na?cæg* ‘hole’ [G99: 14]
- Voiced stops occur intervocalically
  - e. V\_V e.g. *pibe?* \**pipe?* ‘foot’ [G99: 39]

The claim then is that lenition happens generally across the board accounting for the emergence of voiced stops in stressless positions. To explain the occurrence of voiced stops in stressed positions as well as their contrast with voiceless stops in the same positions, an extra ingredient needs to be added. It is argued that voiced stops are allowed in the input, but only if the syllables that contain them carry stress. Stress is preserved in the output due to top-ranked FAITH STRESS, while feature identity ensures that voicing will also survive through the use of high-ranking MAX-[voi]. A few more constraints are added to complete the picture. Effectively then, the proposal maintains that it is not voicing which is unpredictable in Karo, but rather it is stress.

This analysis offers an accurate account of the distribution of voiced stops (less so of voiceless stops) and implicitly suggests that the location of stress is completely accidental due to its unpredictability. In fairness, this view’s major advantage is that it insightfully accounts for the lenition processes that appear in the language. It should be clear that Blumenfeld’s analysis shares with the current one the reference to the relation between voicing and stress. However, while he claims that stress conditions voicing, I claim the opposite, namely that voicing conditions stress.

Nonetheless, while the ‘stress conditions voicing’ approach may seem at a first glance plausible, I will show that there are several points where it proves inadequate or makes wrong predictions. First, recall from (117) that Blumenfeld’s goal is to explain the distribution of voicing through the absence of contrast between voiced and voiceless stops, unless underlying stress is involved in which case both patterns may emerge. But while this analysis superficially seems to discard underlying distinctions of voicing by rendering it predictable, it actually makes crucial use of them with reference to stressed syllables. In fact, it ties voicing with stress (but not the other way round) in a manner which seems rather arbitrary or at best left without any justification.

The burden thus moves to the input stress specification (rather than contrastive voicing). Under this account, all stops are voiceless in the input unless they are stressed in which case they can also be voiced. On the surface then, the only voiced stops will be those in syllables that bear underlying stress or those which are the product of lenition as described in (117e). However, such an approach restricts the input significantly by posing an underlying limitation, namely that voicing should go hand in hand with stress. Due to the Richness of the Base (Prince and Smolensky 1993/2004) and the alleged predictability of voicing we would need to consider inputs with voiced consonants in stressless positions too, e.g. as in the first consonant in hypothetical /baki/ (in line with Blumenfeld, I assume that lexical stress specification occurs in the input). Under Blumenfeld’s account and without further modifications, this would be bound to emerge as [baki], because of high-ranked MAX-[voi] which bans the loss of [voi] in consonants. But this is wrong, since as we have already seen in (117c), word-initial stops are always voiceless.

This problem however is resolved if we allow underlying voicing contrasts - so as to be consistent with ROTB - and permit separate cross-linguistically justified processes to account for the given distribution. In particular, (117c) could be attributed to word-initial fortition (Lavoie 2001), while strengthening of onset stops in stressed positions (117a) is also a possibility (Lavoie 2001, González 2003), although admittedly, a less common process. (117d) looks like a familiar case of final devoicing, while (117e) also illustrates a widespread process of intervocalic lenition. Note that if Blumenfeld were to consider a fuller range of inputs, as hinted above, processes like the aforementioned ones would one way or another be needed to account for the fortition of /baki/ to [paki].

An additional important point which is merely mentioned but not discussed in Blumenfeld is also the following. Voiceless stops in front of stressed syllables (117a)

get lengthened [Gabas 1998: 10 only voiceless stops have long allophones]. This is significant in two respects; first it supports the idea of fortition, since gemination is a familiar case of strengthening (Lavoie 2001, González 2003). Moreover, it seems to support the idea that such consonants carry moras. Given that the effect of stress attraction appears on the final syllable as in e.g. pakɪɔ, it seems reasonable to suggest that this geminate is actually wholly included in the final syllable, i.e. in the onset, instead of having the more familiar flopped representation of geminates (Hayes 1989) where one part of the geminate is in the coda position of the preceding syllable and the other in the onset of the following one (on the possibility of geminates being treated as moraic onsets, see Chapter 6). If this is on the right track, then it just seems that abstract weight is enhanced by lengthening. The gemination pattern is totally missed in Blumenfeld<sup>46</sup>.

It is additionally unclear whether words with - default - final stress should have any underlying specification of stress or not (although in /baki/ above I assumed that they do for illustration purposes). Moreover, claiming that Karo has a partially unpredictable stress system groups it among other languages with lexical stress systems such as Russian or Greek (Revithiadou 1998, Alderete 1999). In these languages however, it has been convincingly shown that morphemes (roots and affixes) bear particular stress properties or requirements in the input. This occurs in a constrained and consistent manner, unlike Karo, where the situation seems to be random.

The most important defect however of Blumenfeld's analysis is one that he also recognises. If it were the case that stress was lexically specified, then we should expect unpredictable stress placement sometimes on the penult and sometimes on the ultima irrespective of whether the final syllable had a voiced onset or not. And yet, penult stress only occurs when the final onset has a voiced onset (setting aside of course the cases where the prevalence of tone or nasalization commands otherwise). More concretely, apart from a handful of exceptions, there are no instances where the final onset has a sonorant or a voiceless obstruent and stress appears on the penult. These cases consistently present final stress.

---

<sup>46</sup> However, in examples such as: [naʔtupʰ] or [piʔti] no lengthening applies. If the geminates are wholly syllabified in the ultima onset, as suggested, at first glance no straightforward explanation is available as to why \*[piʔ.tti] fails to arise. On the other hand, if gemination had the usual flopped representation, i.e. as in \*[piʔt.ti], then the form would be easily eliminated given that complex codas are banned. However, it is more likely that this is more of a phonetic effect. Having a ʔC: sequence would be very difficult to produce as it would require sustaining stop closure for a very long time. Moreover, given that the coda that arises is a glottal stop, some amount of its glottal closure would end up with the following consonant, but even clusters of C'.C are commonly not desirable, e.g. in Tonkawa (Kisseberth 1970).

This is entirely unexpected in an ‘unpredictable stress’ account. Blumenfeld does not solve this problem, but merely states that it is one shared with the way Gabas would have to analyse the data. In particular, Gabas also acknowledges the existence of exceptions, e.g. *kɔɾɛ<sup>b</sup>m* and not *\*kɔɾɛ<sup>b</sup>m* ‘also’ (G99: 41)<sup>47</sup>, so he would need to treat these as cases of underlyingly specified stress and use undominated FAITH STRESS to explain them. But once FAITH STRESS is used, then nothing precludes consideration of inputs such as /*mani*/ where the penult is stressed in the input although the ultima has a non-voiced onset. The expectation is now that this should surface as [*mani*] even though empirically this does not occur.

On the whole then, Blumenfeld’s analysis presents numerous shortcomings and is less economical than the current proposal. To get the Karo data right, Blumenfeld needs to assume lexical specification of stress, underlying voicing (despite his claim that this is not needed) and a close tie between voicing and stress in the lexicon. On the other hand, the present proposal only makes use of underlying voicing distinctions and generates stress on the surface in a predictable manner.

For all these reasons, it seems to me that the proposal outlined in the previous sections is more advantageous than the ‘unpredictable stress’ analysis. In addition, it proves to be more successful in terms of how many correct patterns it predicts, as illustrated in (118). Although a couple of patterns are missed, these could either be superficial exceptions, or could be attributed to various other reasons instead of being ascribed to the structure of the grammar (see below). Comparing the results for each of the analyses will make this clear.

---

<sup>47</sup> In my analysis, these are not exceptions (cf. §2.4.3.2.2).



- (118) Comparison between Blumenfeld's 'lexical stress' analysis and current analysis  
[a check mark denotes that the prediction is correct given the data; italics denote that the prediction is not entirely correct]

	<i>Combination</i>		<i>Attested</i>	<i>Predicted</i>	
				<i>Blumenfeld</i>	<i>current work</i>
a.	+voi	+voi	kiribɔp', miririɣ	✓ YES	✓ YES (cf. (105a))
b.	son	son	wayɔ, mani	✓ YES	✓ YES (cf. (103a))
c.	+voi	son	ayaya, paramit'	✓ YES	✓ YES (cf. (102a))
d.	son	+voi	waro, n <sup>d</sup> ɔgat', yaβa	✓ YES	✓ YES (cf. (101a))
e.	-voi	+voi	cɛgo, parat', karo	✓ YES	✓ YES (cf. (101a))
f.	+voi	-voi	ɛɛpɔ	✓ YES	✓ YES (cf. (102a))
g.	son	-voi	na?tup', mociɔy, nahek'	✓ YES	✓ YES (cf. (103a))
h.	-voi	-voi	NO	✓ NO	✓ NO (cf. (103b))
i.	+voi	-voi	NO	✓ NO	✓ NO (cf. (102b))
j.	son	-voi	NO	✓ NO	✓ NO (cf. (103b))
k.	+voi	+voi	NO	YES	✓ NO (cf. (105b))
l.	son	son	NO	YES	✓ NO (cf. (103b))
m.	+voi	son	NO	YES	✓ NO (cf. (102b))
n.	son	+voi	NO	YES	✓ NO (cf. (101b))
o.	-voi	son	NO	YES	✓ NO (cf. (103b))
p.	-voi	+voi	acibe?, pɔbɔ, cagɔp'	✓ YES	NO (cf. (101b))
q.	-voi	son	pewit'	✓ YES in [σσ] <sub>wd</sub> ; otherwise lenition	YES generally (cf. (103a))
r.	-voi	-voi	pakɔ, ca?pe, pi?ti	✓ YES in [σσ] <sub>wd</sub> ; otherwise lenition	YES generally (cf. (103a))

Glosses: *a.* kiribɔp' 'frog(sp.)', miririɣ 'toad (sp.)' [G99: 41]; *b.* wayɔ 'jacaré' [G98: 40], mani 'macaxeira' [G98: 24]; *c.* ayaya 'cacau' [G98: 20], paramit' 'aranha' [G98: 24]; *d.* waro 'caracol' [G98: 26], n<sup>d</sup>ɔgat' 'comer (intr)' [G98: 16], yaβa 'paca' [G98: 18]; *e.* cɛgo 'macaco (esp.)' [G98: 25], parat' 'curimba (peixe)' [G98: 40], karo 'arara' [G98: 49]; *f.* ɛɛpɔ 'tamanduá' [G98: 25]; *g.* na?tup' 'final' [G98: 27], mociɔy 'mucura' [G98: 27], nahek' 'moleira' [G98: 38]; *h.* acibe? 'raiz' [G98: 76], pɔbɔ 'owl' [G99: 41], cagɔp' 'dish' [G99: 41]; *q.* pewit' 'honey' [G99: 21]<sup>48</sup>; *r.* pakɔ 'pacu' [G98: 40], ca?pe 'escama' [G98: 25], pi?ti 'pesado' [G98: 24]

This table can be split in four parts. In the first, both analyses agree in correctly predicting certain examples (118a-g) as attested; in the second, they do the same but for unattested cases (118h-j). The third part, i.e. (118k-o) is particularly important, since it highlights those instances which only the present approach correctly predicts as unattested, while Blumenfeld fails to do so. Finally, (118q, r) present two cases where the right predictions are made half of the time under the current account, and only one case (118p) which is correctly produced in Blumenfeld's account, but not in the present one. It is thus evident from the above that Blumenfeld's approach captures 13/18 patterns, while the current approach captures 15/18.

<sup>48</sup> Note that Gabas (1998: 51) transcribes this as *pewít'* instead, i.e. with a high tone which would already attract stress.

The analysis that utilizes lexical stress specification predicts numerous cases for which no data exist. This is because it fails to see the connection between the moraicity and voicing properties of the segments. A few cases however are correctly ruled out (118h, i, j), not because of stress considerations, but because they should be impossible forms. In particular, they involve voiceless obstruents in stressless non-initial positions<sup>49</sup>, which according to Blumenfeld, should be lenited and appear with voiced obstruents instead. Forms like (118q) and (118r) are predicted by Blumenfeld, but as I note down, only in disyllabic words. This is because the first -voi segment will remain -voi word-initially as in a disyllabic word, but will turn into +voi once it finds itself word-medially due to lenition, as would happen in a trisyllabic word, e.g.  $\sigma_{[-voi]} \sigma_{[-voi]} \sigma_{[-voi]} \rightarrow \sigma_{[-voi]} \sigma_{[+voi]} \sigma_{[-voi]}$ . Examples of this form exist, e.g. [paramitʰ] ‘aranha’ (G98: 11), [ciβɛkɪɓ<sup>d</sup>n] ‘urubu’ (G98: 18). The present approach makes - possibly wrongly - no similar distinction and predicts that -voi followed by a stressed syllable with a -voi/son onset should occur independently of the length of the word. On the other hand, note that (118k) is wrongly predicted in Blumenfeld’s account because of the unpredictability of stress. For the same reason, we find penultimate stress, where we should not, as in (118l, m, n, o).

Despite a few shortcomings, the current analysis accounts for the empirical facts more adequately than its alternative. There are only a few words as in (118p) which disrespect the stress shift pattern, but I contend that these are far too few to weaken the generalisation. Moreover, instead of imposing lexical stress specification and faithfulness to underlying stress for exceptions, it seems to me that it is equally likely or even plausible to consider that some other factor may be involved here. For instance, as we have seen, high tone or nasalization outweigh conditions of stress placement based on onset quality, so it could perhaps be the case that such words actually have a H-tone final syllable and this has simply not been recognised, e.g. cf. the case of *pewitʰ* (118q and fn. 48). It would also be worthwhile looking at the general composition of the lexicon, e.g. whether there are any loanwords and if so, to what extent these affect the grammar.

Given the current state of affairs and the data available, the present approach has examined the facts as fully as possible and has satisfactorily accounted for Karo default

---

<sup>49</sup> Word-initially, syllables with stressless voiceless obstruent onsets are protected, while word-medially they appear voiced. This is a fact that Blumenfeld captures through lenition processes, whereas the current analysis, fails to do so, but could be amended accordingly. This is why (118q, r) are only partially correctly predicted, i.e. there are no finally stressed trisyllables where  $\sigma_2=[-voi]_{On}$  and  $\sigma_3=[-voi]_{On}$  or where  $\sigma_2=[-voi]_{On}$  and  $\sigma_3=[son]_{On}$ .

stress and stress shift. At the same time it has attempted to minimize the set of the assumptions involved. It thus proves to be advantageous over Blumenfeld's alternative analysis both empirically and conceptually.

## 2.5 Conclusion

In this chapter we have studied three languages in detail: Pirahã, Arabela and Karo. All of them present onset effects on stress, which I have analysed by making reference to the moraicity of onsets. Pirahã's stress algorithm is sensitive to both the presence as well as the quality of an onset. Arabela pays attention to the type of onsets and under certain circumstances undergoes stress shift as a response to that. Both languages share the fact that they only treat voiceless obstruent onsets as moraic, leaving sonorant and voiced obstruent ones moraleless. Finally, Karo also presents a stress shift pattern based on the type of onsets involved, but differentiates itself in that it treats sonorants as moraic instead of non-moraic.

These onset moraicity patterns are easily accounted for by the ranking  $*\mu/\text{ONSET}/[\text{voice}] \gg \text{MORAIC} \gg * \mu/\text{ONSET}$  and the presence of [voice] for sonorants (Pirahã, Arabela) or its absence (Karo). The general WSP constraint proves invaluable in all the cases as it forces stress on heavy syllables due to moraic onsets. Its interaction with other constraints in the languages at hand yields stress assignment or stress shift facts. Onset moraicity allows us to account for these facts in a uniform way. Note however that stress on its own cannot support the onset moraicity hypothesis adequately. The reason is - as thoroughly explained in the previous chapter - that stress is subject to a re-analysis along the lines of prominence-based accounts. We thus need more robust evidence that comes from phenomena which cannot possibly receive any prominence explanation. Obviously this refers to weight-based phenomena sensitive to onsets. Chapters 4-7 deal exactly with this sort of cases that include word minimality, compensatory lengthening, geminates and reduplication.

However, before exploring these phenomena, I will address the importance of the *presence* of an onset in stress systems. As I have claimed before, and will argue next more extensively, this does not bear on the issue of moraic onsets as it is a separate phenomenon. Nonetheless, since it offers a better understanding of onset-sensitive stress and since some of the Pirahã facts pertain to it, I will devote the next chapter to the 'onset presence' effects for completeness.

## Chapter 3

### Onset and stress effects due to the presence of an onset

#### 3.1 Introduction

Up to now, I have been examining languages where the effect of onsets on stress depended on their quality, so that certain onsets attracted stress compared to others which did not. This was attributed to onset moraicity considerations. However, recall from §1.2 that stress could also be affected by merely the presence of an onset. There are therefore two factors relating to onset sensitivity and stress: i) the quality of the onset, and ii) the presence of the onset. In this section I will focus on the latter.

Since I argue that the effect the presence or absence of an onset may have on stress is a different phenomenon from that relating to the type of the onsets involved, we should expect that both phenomena could simultaneously emerge in a single language. This expectation is borne out in pattern ❶, as shown in (1) in the case of Pirahã (§2.2). Recall that in Pirahã, apart from the fact that PV(V) syllables attract stress more than BV(V) ones by virtue of onset moraicity, it is also the case that BVV > VV. This means that among syllables of equal moraicity, the onsetful ones are preferred to receive stress.

(1) *Presence and quality of onset interaction in stress*

		Quality of onset	
		YES	NO
Presence of an onset	YES	Pirahã (❶)	Aranda, Banawá, Dutch (❷)
	NO	Karo (❸)	Standard Greek, Russian, etc. (❹)

In section §2.2.2 devoted to Pirahã, I had used the constraint ALIGN<sub>σ</sub>O to account for this fact. Here, I will discuss this constraint in more detail, justify its use and consider its effects with respect to several other languages, namely the ones under the cell in ❷ above. I will argue that ALIGN<sub>σ</sub>O is the driving force for two phenomena: i) the retraction of stress to an onsetful syllable as happens in Aranda, Banawá and Pirahã [i.e. an onset-on-stress effect (OSE)] and ii) the creation of an onset in a stressed syllable, as happens in Dutch and marginally in English [i.e. a stress-on-onset effect (SOE)].

### 3.2 Justifying ALIGN<sub>FO</sub>

The Pirahã weight and stress system includes the following five-way weight hierarchy: PVV > BVV > VV > PV > BV. This system has been extensively explored in §2.2.2, but the BVV > VV part of this scale, highlighted in (2), will serve as the starting point of the discussion in this section. Pirahã stress is located on one of the last three syllables in the word. Default stress is final, as becomes evident when looking at words that contain syllables of the same nucleus and onset structure (2a).

(2) *Pirahã default stress (a); Pirahã BVV > VV part of the scale (b)*

a.	ko.ʔo.pa	“stomach”	[E: 239]
	bíi.gao.baá	“certainly called”	[E: 239]
b.	pòò.gáí.hì.àí	“banana”	[E&E: 709]

Examples like the one in (2b) show is that all else being equal, an onsetful syllable is preferred to receive stress over one that lacks an onset, i.e. gai > ai. Thus the onset of the syllable must trigger this stress shift. One way to capture this effect is by using the constraint ALIGN<sub>FO</sub> which is an updated version of Goedemans’ (1996, 1998) ALIGN-FTO constraint.

- (3) ALIGN<sub>FO</sub>: Align-L (σ, C), i.e. Align the left edge of every stressed syllable with a consonant
- (4) ALIGN-FTO: Align-L (Ft, C), i.e. Align the left edge of every foot with a consonant<sup>1</sup>

There are some reasons why (3) should be preferred over (4). What (4) says is that the left edge of the foot has to start with a consonant. Although Goedemans does not discuss this issue, this constraint essentially accounts for trochaic languages, but not for iambic ones, as a result of the fact that ALIGN-FTO merely targets the left edge of a foot irrespective of its type. To see why this is the case, let us consider the following cases: a bisyllabic trochaic foot (σσ) and a bisyllabic iambic foot (σσ̄). In the first case, the left edge of the foot coincides with the left edge of the foot-head, i.e. the stressed syllable.

<sup>1</sup> This can be seen as a positional markedness constraint that favours the alignment of feet with consonants (Zoll 1998).

In the latter though, the left edge of the foot happens to be the left edge of the foot-tail, since the foot-head lies on the second syllable. Therefore, such a constraint makes the prediction that in an iambic system the onset effect will not appear on the stressed syllable, but on the stressless one. This seems a highly unlikely and potentially problematic prediction.

Perhaps, Juma - to the extent that the data described according to Abrahamson and Abrahamson (1984) and mentioned below are accurate - offers a concrete example. In Juma, the final syllable gets stressed unless it is a V in which case stress moves to the penult. Whether the foot is iambic or trochaic cannot be determined, but this shift would only work if the foot is trochaic, i.e. in final stress we would have a degenerate foot as in  $\dots\sigma(\acute{\sigma})\#$  and then with shift, a binary  $\dots(\acute{\sigma}\sigma)\#$  with the foot-edge coinciding with an onsetful syllable. If the foot is iambic however, default final stress would work fine<sup>2</sup>, i.e.  $\dots(\sigma\acute{\sigma})\#$  but stress shift to the penult would produce  $\dots(\sigma\acute{\sigma})\sigma\#$ . The problem is now that ALIGN-FTO cannot account for this shift, since the left edge of the foot does not coincide with the stressed syllable which has to be onsetful. ALIGN $\acute{\sigma}$ O however does not face this problem, since it directly refers to the stressed syllable.

Another peculiar prediction of ALIGN-FTO is the following. Consider a language which forms iambs from left to right, e.g.  $(CV.\acute{CV}).CV.CV$ <sup>3</sup>. Suppose further that this language shows the onset-sensitivity effects described in this section. Now, take the string  $V.CV.CV.CV$ . How would this be parsed? According to ALIGN $\acute{\sigma}$ O, the left edge of the stressed syllable must begin with a consonant yielding the parsing:  $(V.\acute{CV}).CV.CV$ . ALIGN-FTO though will select the parsing where the first syllable is left unparsed, i.e.  $V.(CV.\acute{CV}).CV$ . Essentially under ALIGN-FTO, the stress algorithm for this language would say 'stress the second syllable if the first is onsetful, but stress the third if it is onsetless'. On the other hand, ALIGN $\acute{\sigma}$ O would consistently build iambs without such a seemingly poorly-grounded stress shift.

Despite these advantages of ALIGN $\acute{\sigma}$ O, at this stage, it seems rather inconclusive to determine which formulation should be chosen, given that to my knowledge there is no clearly iambic system exhibiting such effects relating to stress. However, according to Green's (1995) analysis, Pirahã is a mixed system where both trochees [for antepenultimate and penultimate stress] and iambs [for final stress] are employed. If this is right, only ALIGN $\acute{\sigma}$ O can be used to yield the correct results for both trochaic and

<sup>2</sup> Assuming of course that the penult is onsetful.

<sup>3</sup> I ignore here the strong tendency for iambs to have a heavy syllable in the foot-head and make no reference to additional feet that could be built in these strings, as they are not relevant to the argument.

iambic patterns. But even so, the fact that ALIGN $\sigma$ O does not in the first place refer to any footing, can be advantageous in cases where analyses agree on where stress is located, but not quite on the exact footing (such as Juma or Dutch below).

One might wonder if a possible advantage of ALIGN-FtO over ALIGN $\sigma$ O is that the former constraint resembles the formulation of the fundamental constraint ONSET rephrased in terms of alignment by McCarthy and Prince (1993) and Itô and Mester (1994), as shown below.

(5) ONSET: Align-L ( $\sigma$ , C), i.e. Align the left edge of a syllable with a consonant

This is similar to the formulation of ALIGN-FtO, i.e. Align-L (Ft, C), the only difference being that the argument of the prosodic constituent has changed from a syllable to a foot. ALIGN $\sigma$ O's definition is repeated below.

(6) ALIGN $\sigma$ O: Align-L ( $\sigma$ , C)

Again this is identical to the formulation of ONSET, but what changes this time is that the syllable has to be stressed. This may be seen as not particularly plausible, given that direct reference to stressed syllables is made, but this is not necessarily so, since it is well-known that stressed syllables are prominent positions. In the same vein of constraints like Anchor- $\sigma$  utilized by Nelson (1998) for French truncation and Max- $\sigma$  as in Madsen (2000) for Spanish blend formation or Beckman (1999) for Scots Gaelic, alignment constraints may be able to refer to such positions. Thus, given the existence of several constraints specific to stressed positions and the potentially hazardous predictions of ALIGN-FtO, the use of ALIGN $\sigma$ O throughout is justified.

Note that in this view, ALIGN $\sigma$ O is virtually identical to Smith's (2005) ONSET/ $\sigma$  constraint, which requires that a prominent position, like a stressed syllable, has an onset. There are two reasons however why I do not choose this formulation. By using ONSET/ $\sigma$  as the relevant constraint, one might assume that I espouse Smith's theory as a whole. Nonetheless, as I have shown in §1.3.2, there are some difficulties in this model, particularly when stress is sensitive to the quality of onsets, rendering it inadequate to explain this phenomenon satisfactorily.

The second reason is the following: if alignment is the right way to think of sensitivity to the presence of an onset, then under an analysis along the lines of

ALIGN $\phi$ O (or for that matter, ALIGN-FTO too), we could expect to find a corresponding constraint at the right edge, so that the right edge of a stressed syllable is closed by a consonant. This is not however possible in Smith (2005), since it would imply the existence of a constraint CODA/ $\phi$ , which is not predicted by the positional augmentation model she puts forward. And yet a potential example of this kind seems to appear in the Brazilian language of Kaingang (Yip 1992), a language lacking long vowels and geminates, but allowing the nasal codas *m*, *n*, *ɲ*, *ŋ* with their allophones and more controversially the voiced continuants *w*, *y*, *r*. What is interesting for our purposes is that under certain conditions, stress-attracting penultimate open syllables acquire a default velar nasal in reduplication - and sometimes outside reduplication too - so that the syllable becomes heavy and thus can more easily accommodate stress. A constraint like ALIGN-R ( $\phi$ , C) would straightforwardly account for this fact.

It is also conceivable to think of Italian Raddoppiamento Sintattico in a similar way. Basically what happens there is that in a sequence of two words  $W_1$  and  $W_2$  where the last syllable of  $W_1$  ends in a stressed vowel and  $W_2$  begins with a consonant, the initial consonant of  $W_2$  geminates providing a coda for the final syllable of  $W_1$  (Borrelli 2002, McCrary 2002).

### 3.3 Patterns that ALIGN $\phi$ O triggers

ALIGN $\phi$ O states that a stressed syllable must be onsetful, but does not impose any requirements on how this can be achieved. There are in fact at least two ways in which ALIGN $\phi$ O can be satisfied. The first involves manipulation of the general stress system of the language so that stress falls on an onsetful syllable. In this instance, the presence of an onset affects stress placement by means of stress shift, rendering such shift an onset-on-stress effect (OSE). Another way to satisfy ALIGN $\phi$ O is by manipulating the segmental structure. In other words, stress assignment applies as usual, but if it happens to be the case that it would fall on an onsetless syllable, then an onset is inserted. This constitutes a stress-on-onset effect (SOE) because it is stress that triggers onset epenthesis. Both patterns are attested as shown below. I will briefly outline the patterns we find and then discuss them in more detail in the forthcoming subsections.



(I) *OSE effects*

In languages like Aranda, Alyawarra, Banawá or Iowa-Oto, the stress algorithm is determined based on ALIGN<sub>OO</sub>. Stress is normally placed on the first syllable, unless this is onsetless, in which case it shifts to the second (onsetful) syllable. Here the relevant ranking is: ALIGN<sub>OO</sub> >> ALIGN-L (PRWD, FT)

(II) *SOE effects*

In languages like Dutch, the general stress algorithm assigns stress on a syllable, which, if onsetless, acquires an onset, e.g. Dutch: xá.ɔs but a.ʔór.ta. The interesting property is that the inserted consonant is /ʔ/. This makes sense for various reasons, one of which is argued in this thesis. It is non-sonorous, it is unmarked in terms of place specification, and by being voiceless, it increases the pitch of the following vowel providing a cue that it is stressed. The general ranking that holds here is: ALIGN<sub>OO</sub> >> DEP-C

### 3.3.1 Onset-on-Stress Effects (OSE)

#### 3.3.1.1 An overview: Aranda, Alyawarra, Banawá and the role of ALIGN<sub>OO</sub>

The OSE effect is illustrated by a handful of languages, especially Australian ones, which present the following, roughly stated, stress algorithm.

- (7) Stress the first syllable if onsetful, otherwise stress the second one.

Languages which have this algorithm include: Aranda (Strehlow 1944), Alyawarra (Yallop 1977) and other Australian languages, such as Lamalama, Mbabaram, Umbuygamu, Linngithig, Kuku-Thaypan, Kaytetj and Agwamin (most of them are Cape York and Arandic ones; see Goedemans 1998 and Blevins 2001 for more details). Also the native American Iowa-Oto (Robinson 1975), Banawá (Buller, Buller and Everett 1993), and Juma (Abrahamson and Abrahamson 1984). The latter presents this effect at the right edge of the word, so that the final syllable is stressed if it is a CV one, but if it is a V one, then the penult gets stress instead. I start with examples from Aranda.

(8) *Aranda stress*

a. Consonant-initial words of 3 or more syllables

rá:tama	'to emerge'
kútungùla	'ceremonial assistant'
lélantinama	'to walk along'

b. Vowel-initial words of 3 or more syllables

ergúma	'to seize'
arálkama	'to yawn'
ulámbulàmba	'water-fowl'

c. Words of 2 syllables (C- or V- initial)

ílba	'ear'
á:twá	'man'
kála	'already'
gúra	'bandicoot'

In Aranda, the algorithm presented in (7) is in operation, so that C-initial words receive stress on the first syllable (8a), but V-initial onsetless ones, have stress on the second syllable (8b). There is however an exception to this pattern, namely that in disyllabic words, stress is word-initial independently of whether the word begins with a vowel or a consonant. Presumably, this relates to the fact that Aranda avoids final stress due to NONFINALITY (or prefers to create binary feet according to Goedemans (1996)), thus words like \**il(bá)* or \**a(rálka)(má)* with final stress are impossible.

The closely related Alyawarra has a similar pattern, with two exceptions. First, syllables beginning with a glide pattern with onsetless syllables<sup>4</sup>. Second, disyllables show the same behaviour as polysyllables do. This is apparently because in Alyawarra, non-finality considerations submit to the need to assign stress on onsetful syllables. When this is not at stake however, final stress is still avoided e.g. \**arrákirtá*.

---

<sup>4</sup> This is not as surprising as it seems. Smith (2003) argues that there are two types of glides; the true onset ones, and the nuclear ones. The latter are actually part of the nucleic node and thus exempt from statements that refer to onsets. Smith claims that among related Campidanian Sardinian dialects, the Iglesias dialect has nuclear onglides, whereas the Sestu dialect has the true onset ones. In addition, there are cases where the distinction between the two glides is found in a single language as is the case in French, Spanish, Slovak and English. For details, the reader is referred to Smith (2003). The implication of this result is that in Alyawarra, onset glides could be treated as nuclear onglides, which means that they are not true onsets. Consequently, the resemblance between glide-initial and onsetless syllables falls out naturally. Moreover, this theoretical assumption is compatible with Yallop's statements that "glides are part of the phonetic realisation of the vowels" (1977: 19), as well as that "wa- and yu- word-initially represent phonologically simple vowels" (1977: 20).

(9) *Alyawarra stress*

a. Consonant-initial words of 2 or more syllables

kwíya	‘girl’
kíra	‘meat’
párrí:ka	‘fence’
ngáyakwa	‘hungry’

b. Vowel-initial or glide-initial words of 2 or more syllables

athá	‘I (ergative)’
i:lpá	‘ear’
ilípa	‘axe’
arrákirta	‘mouth’
walí:mparra	‘pelican’

In Banawá, feet are built on moras (not on syllables) and vowel sequences form tautosyllabic diphthongs. Long vowels do not occur. Banawá is quite similar to Aranda in that it too stresses the onsetless syllable in bimoraic disyllabic words (10c), because otherwise word minimality, which requires a bimoraic foot per word, would be violated. However, in words of increased moraicity, the avoidance of stressed onsetless syllables can freely operate, which is why stress shifts to the second syllable (10b). In consonant-initial words, the algorithm operates as normal, assigning stress on the first syllable, since it is onsetful.

(10) *Banawá stress*

a. Consonant-initial words

dísa	‘shoot with an arrow’
bádi	‘name’
mákarì	‘cloth’
fúanà	‘lost’
tátikùne	‘hair’
ùnarífabùne	‘you are going to work’

b. Vowel-initial words of more than 2 moras

idífa	‘to marry’
owárià	‘one’
ufábunè	‘I drink’

c. Disyllabic vowel-initial words of 2 moras

ába	‘fish’
íta	‘sit’
áwi	‘tapir’

Finally, in Juma, data are limited, but again, stress seems to be sensitive to the presence of an onset. Juma differs from all the languages above in that this effect emerges word finally, i.e. *ǝkúǝ* ‘esta cortando’ where the final syllable is onsetful, versus *peyǝkǝpía*

‘espécie de passaro’ where it is not. However, a word of caution is in order. In the limited data available, it is possible to re-analyze cases like *peyíkopia*, as containing a final diphthong, i.e. *pe.yí.kɔ.pía* rather than a sequence of heterosyllabic vowels, i.e. *pe.yí.kɔ.pl.a*, which is what Abrahamson and Abrahamson (1984) seem to suggest. If the diphthong analysis proves right, then Juma no longer seems to illustrate onset-sensitivity of this type.

It should be obvious that all these data share one thing in common: Stress preferably docks on the first onsetful syllable (almost always counting from the left edge of the word rightward).

A question that comes to mind is whether the facts above can be analysed in a way other than using ALIGN<sub>σ</sub>O. One alternative is the one offered by Downing (1998) who observes that in some languages onsetless syllables are treated as ill-formed due to the lack of onset and consequently are excluded from certain prosodic processes such as stress, tone and reduplication. In the default situation, she explains, morphological and prosodic constituents are co-extensive with one another. However, there are also constraints which require that prosodic constituents are aligned with optimal, i.e. onsetful syllables. If these constraints are high-ranked, misalignment between morphological and prosodic constituents is enforced, leaving the ill-formed onsetless syllables outside the prosodic domains, i.e. they are extraprosodic. As a result, onsetless syllables do not participate in the prosodic processes in question, hence their exceptional behaviour.

While this account has good empirical coverage particularly with respect to tone and reduplication, it fails to account for some instances of onset-sensitive stress. In particular, since it heavily relies on matching morphological constituents with prosodic ones, it is designed to act at the edges of morphemes and prosodic constituents. Thus, although it can account for the cases presented in this section, since they refer to the (left) edges of words, it fails to capture the facts in Pirahã where the BVV > VV hierarchy holds word-medially and intra-morphemically. On the other hand, a constraint like ALIGN<sub>σ</sub>O makes no similar claim, therefore it can freely apply to the case of Pirahã as well.

Of course, it is true that the majority of OSE effects presented here occur at word edges, a fact which is successfully captured in Downing (1998), but is not immediately obvious in the current proposal. One thing to notice with respect to the increased frequency of ALIGN<sub>σ</sub>O effects at the left edge of the word relates to the fact

that most of the languages that present this phenomenon are Australian (especially of the Cape York and Arandic families). These languages are renowned for the historical process of initial dropping, that is, the loss of the initial consonant of the word and stress shift from the first syllable to the second. In this way  $*[ʼCVCV...] > [VʼCV...]$ <sup>5</sup>. This then gives us the ‘stress the first onsetful syllable’ effect, which nonetheless needs to be accounted for somehow synchronically. An account of this sort follows shortly.

Note that there is debate over the series of the events above. Blevins (2001) claims that stress shift occurred first followed by consonant loss<sup>6</sup>. Goedemans (1998) suggests that the order was the reverse, i.e. that consonant-loss served as the trigger of stress shift. Had it been the other way round, he argues, we should find languages where stress shift has already applied, but consonant loss has not yet. Evidence one way or another is not conclusive, and at any rate, orthogonal to the point of interest for us here, namely that - synchronically - onsetful syllables attract stress. More importantly, the diachronic explanation is on its own insufficient, since no evidence is available of a similar process having occurred in the native American languages (Banawá and Iowa-Oto). The use of ALIGN $\sigma$ O is therefore indispensable.

Another reason, of a more functional nature, that could account for the left word-edge preference of ALIGN $\sigma$ O application (with the dubious exception of Juma), relates to well-known properties of certain syllables and positions. Stressed syllables are ‘strong’ positions, and so are word-initial ones (Beckman 1999, Smith 2005, among others). ‘Strong’ positions are those which tolerate marked structures in contrast to ‘weak’ ones which do not. In a theory like Smith (2005; cf. §1.3.2), ‘strong’ positions need to be augmented, i.e. acquire properties that enhance their perceptual salience, such as stress or the presence of an onset. Consequently, a syllable which is word-initial, stressed and onsetful would be maximally salient and hence a desirable configuration across languages. In fact, Smith (2005) suggests the positional augmentation constraint ONSET/ $\sigma$ , which states that if a syllable is stressed, then it must also have an onset. As mentioned above, this constraint is virtually identical to ALIGN $\sigma$ O and can equally account for the facts mentioned below in Aranda or Dutch -

---

<sup>5</sup> Since most of these languages lack complex onsets, this process presumably happened in some, but not all cases, otherwise, all words should have been vowel-initial!

<sup>6</sup> Blevins (2001) reports numerous other Australian languages where she claims the sequence of facts was stress shift → C-loss. These include Mpalitjahn, Yinwum, Linngithigh, Anguthimri, Ngkoth, Mbiywom, Rimanggudinhma, Umbuygamu, Lamalama, Kuku-Thaypan, Mbabaram, Wamin, Yanga, Mbara, and more questionably Ogunyjan, Olgol, Oykangand and Nganyaywana.

see Smith (2005) for detailed analysis. However, it carries along certain shortcomings (cf. §1.3.2 and end of §3.2), which is why I do not adopt it.

In the next section I will show how ALIGN<sub>FO</sub> can be implemented. An account along similar lines has been offered in Goedemans (1996) for Aranda and Alyawarra, which is why I will instead prefer to provide a short analysis of a lesser-known, barely discussed, stress system, that of Iowa-Oto (Robinson 1975)<sup>7</sup>.

### 3.3.1.2 Iowa-Oto

Iowa-Oto is part of the Chiwere sub-group of the Siouan language family. It is practically an extinct language, since according to the Ethnologue, the last fluent speakers died in 1996. A few remain with some knowledge of the language.

The stress system is virtually identical to the one of Alyawarra, although glide-initial words seem to pattern with consonant-initial ones. Consonant-initial words receive primary stress on the first syllable (11a). In vowel-initial words, primary stress docks on the second syllable (11b). The number of syllables in a word is irrelevant.

#### (11) *Iowa-Oto stress* (Robinson 1975)

##### a. Consonant-initial words

péce	‘fire’
náwe	‘leaf’
hérota	‘morning’
páxoče	‘Iowa’
wíwaθòče	‘machine’

##### b. Vowel-initial words

aháta	‘outside’
it <sup>h</sup> á	‘there’

These data obviously suggest that ALIGN<sub>FO</sub> is involved here, favouring stress on onsetful syllables and causing the shift of it, if the first syllable is onsetless, i.e. ALIGN<sub>FO</sub> >> ALIGN-L (PRWD, Ft) [i.e. Align the left edge of the PrWd with the left edge of a Foot; abbreviated as ALIGN-L]. With regard to secondary stress, we can

---

<sup>7</sup> Another choice could be Banawá, which has not been dealt with in terms of ALIGN<sub>FO</sub> or ALIGN-FTO, but has been examined by Buller, Buller and Everett (1993), Everett (1996), Downing (1998) and Gordon (2005), all of which are however quite different analyses from the present one. I contend that Banawá too can be translated in terms of ALIGN<sub>FO</sub>. However, since it presents some complications regarding the vowel combinations it allows, minimality considerations and foot parsing, providing a full account would take us too far afield. Note that Aranda is also analysed in Takahashi (1994), Downing (1998) and Smith (2005), while Alyawarra is also dealt with in Downing (1998). To my knowledge, Iowa-Oto is only discussed in passing by Gordon (2005), so the present analysis is more fully-fledged.

observe that while rhythmic stress is desirable as in *wíwaθòčē*, it is refrained from docking on the final syllable, e.g. *hérota* not *\*hérotà*. This can be attributed to a relatively high-ranked NONFIN or to FTBIN which requires that feet are bisyllabic. While any of these would do, for current purposes, I will use FTBIN so that I also draw attention to the rhythmic nature of stress assignment that prefers binary groupings of feet. What this entails is that syllables may remain unparsed if this rescues them from a FTBIN violation, therefore: FTBIN >> PARSE-σ. Despite that, it is not the case that feet are never unary. They can be, in case stressing an onsetless syllable is at stake as in *it<sup>h</sup>á* not *\*íta<sup>8</sup>*, entailing that ALIGN<sub>σO</sub> is very highly ranked in the language and dominates FTBIN. The following cases illustrate.

(12) ALIGN<sub>σO</sub>, DEP-C >> ALIGN-L

	nawe	ALIGN <sub>σO</sub>	DEP-C	ALIGN-L
☞ a.	náwe			
b.	nawé			*!
	ahata	ALIGN <sub>σO</sub>	DEP-C	ALIGN-L
c.	áhata	*!		
d.	?áhata		*!	
☞ e.	aháta			*

The example /ahata/ → [aháta] compared to the normal stress pattern as illustrated by /nawe/ → [náwe] reveals that ALIGN<sub>σO</sub> >> ALIGN-L. In other words, the regular stress algorithm, which requires primary stress on the first syllable, is disrupted by the presence of ALIGN<sub>σO</sub> which forces shift of stress to the peninitial. ALIGN<sub>σO</sub> could also be satisfied by consonant insertion and stress on the first syllable as in (12d). This is ruled out though because DEP-C is top-ranked too. Consequently, ALIGN<sub>σO</sub> is satisfied by altering the preferred prosodic alignment of the stressed syllable.

<sup>8</sup> Robinson (1975: 443) states that /p t k ʔ/ are aspirated when in a stressed syllable, therefore the /t/ in *it<sup>h</sup>á* would not be aspirated had it not been stressed. However, Robinson is not consistent in her transcriptions, since consonants that would fulfil the aspiration environment are not presented as such, e.g. *péčē* instead of anticipated *p<sup>h</sup>éčē*. For clarity however, I will follow Robinson's transcriptions throughout.

(13) FTBIN >> PARSE-σ

herota	FTBIN	PARSE-σ
a. (héro)ta		*
b. (héro)(tà)	*!	
wiwaθoče	FTBIN	PARSE-σ
c. (wíwa)(θòče)		
d. (wíwa)θoče		**
e. (wíwa)(θò)(čè)	**	

(13a) and (13b) provide the next ranking argument, namely that FTBIN >> PARSE-σ, since disyllabic feet are preferred even at the expense of leaving certain syllables unparsed. In *wíwaθòče*, perfect satisfaction of both constraints is possible as shown in (13c).

The final point arises when we consider a bisyllabic onsetless word. Our grammar yet does not tell us what will happen, since there are two independent conflicting demands. On the one hand, stress needs to dock on an onsetful syllable due to high-ranking ALIGNσO, therefore for an input like /ita/ we would expect [i(t<sup>h</sup>á)]. On the other hand, feet need to be disyllabic due to high-ranking FTBIN, consequently, we would predict [(í<sup>h</sup>a)]. In reality, what we get is [i(t<sup>h</sup>á)], thus we have a new ranking argument, namely that ALIGNσO >> FTBIN, exemplified below.

(14) ALIGNσO >> FTBIN

/ita/	ALIGNσO	FTBIN
a. i(t <sup>h</sup> á)		*
b. (í <sup>h</sup> a)	*!	

All in all then, the Iowa-Oto stress system can be described by the ranking:

(15) ALIGNσO, DEP-C >> ALIGN-L (PRWD, Ft), FTBIN >> PARSE-σ

### 3.3.1.3 Interim summary

The preceding discussion examined the OSE effect, where an onsetful syllable attracts stress due to ALIGNσO. This phenomenon usually occurs at the left edge of the word, but can also appear word-medially, as in Pirahã. Satisfaction of ALIGNσO comes through prosodic misalignment of the normal stress algorithm of the language. For



instance, in Iowa-Oto, stress is initial, unless the word begins with a vowel, in which case it moves to the next syllable producing misalignment to the left-edge of the word, where the preferred position of stress is located. ALIGN<sub>σ</sub>O must therefore dominate the ALIGN-STRESS constraints, and also rank at least as highly as DEP-C. This is important, because had DEP-C been low-ranked, it would be possible to satisfy ALIGN<sub>σ</sub>O not by misalignment, but by the segmental fix the insertion of an onset in front of a stressed syllable could offer. Such cases also exist, as I will show immediately below in the discussion of Dutch.

### 3.3.2 Stress-on-Onset effects (SOE): the case of Dutch

Stress-on-onset effects (SOE) are also driven by ALIGN<sub>σ</sub>O as mentioned previously. Unlike the OSE effects above, here ALIGN<sub>σ</sub>O does not impose any changes to the stress algorithm of the language in question by shifting stress to an onsetful syllable. What it does instead is that if it so happens that stress falls on an onsetless syllable, then this - under certain conditions - acquires an onset.

This pattern marginally appears in British English where words like *co-operate*, *geometry* and *reaction* can be realized in careful speech as [kəʊʔɒpəreɪt], [dʒiʔɒmətri] and [ri:ʔækʃən] respectively (inserted ʔ is underlined; Cruttenden 2001: 169) or when contrasts like [dʒiʔɒgrəfi] ‘geography’ vs. [dʒi:əʊgræfɪk] ‘geographic’ optionally emerge (Michael Ashby, p.c.). While this pattern is facilitated by the presence of stress on the second vowel, it is reported that it can also arise in other hiatus contexts in the absence of stress (Cruttenden 2001: 169). However, even in the cases where stress is assigned on the second vowel, the pattern is not systematic. For this reason, I will instead use data from Dutch, where the phenomenon is robustly documented.

#### 3.3.2.1 The hiatus data when $V_1$ is /a/ and the role of ALIGN<sub>σ</sub>O

The relevant data are presented below.

(16) Dutch *ʔ*-insertion under hiatus when  $V_1$  = /a/ (Booij 1995: 65)

a.	<i>ʔ</i> -insertion		
	/paɛlja/	[pa.ʔɛl.ja]	‘paella’
	/aɔrta/	[a.ʔɔr.ta]	‘aorta’
	/kaunda/	[ka.ʔún.da]	Kaunda [Zambia’s first president]

b.	no ?-insertion		
	/xaos/	[xá:.os]	'chaos'
	/farao/	[fá:.ra.o:] <sup>9</sup>	'Pharaoh'

The data above present a hiatus environment where the first vowel is /a/<sup>10</sup>. In the first set of examples (16a), a glottal stop is inserted between /a/ and the following vowel on the surface. The second set illustrates a near-identical environment, where ?-insertion fails. The crucial difference is that the underlyingly onsetless syllable in (16a) is stressed in the output in contrast to the one in (16b) which is not. We can therefore argue that the reason the glottal is inserted in (16a), but not in (16b) is because ALIGN<sub>OO</sub> is in action, positing a requirement for a stressed syllable to begin with an onset. If ALIGN<sub>OO</sub> >> DEP-C then a consonant is going to be inserted to fulfil this need. No similar requirement however applies for unstressed onsetless syllables, therefore any such insertion would violate DEP-C pointlessly. Note that the inserted consonant, as we see, is the glottal stop /ʔ/, a consonant which is unmarked in terms of its place specification (Prince and Smolensky 1993/2004) in a theory of place markedness where \*LAB, \*DORS >> \*COR >> \*PHAR and where /ʔ/ is treated as a pharyngeal (Alderete et al. 1999, Lombardi 2002). Moreover, by being voiceless, /ʔ/ is entirely compatible with the idea put forward in this thesis, namely that the lack of voicing increases the pitch of the following vowel providing a cue for stress. Here, the stress algorithm alone decides where stress will dock, but if it happens to be on an onsetless syllable, then the addition of /ʔ/ renders such syllable more well-formed by virtue of the presence of an onset. To simplify things - as numerous constraints will be added in due course to capture all details - I will use DEP-ʔ instead of DEP-C.

Before considering an analysis of these facts, let me first give some background on Dutch stress based on Gussenhoven (to appear). This will be quite general, since there are numerous detailed accounts of Dutch stress available which the interested reader can consult. These include van der Hulst (1984), Lahiri and Koreman (1988), Kager (1989), Gussenhoven (to appear) among others. Simplifying a bit, Dutch primary

<sup>9</sup> Carlos Gussenhoven (p.c.) observes that in a word like [fá:.ra.ð:], there is also secondary stress on the final syllable. Insertion of /ʔ/ there is optional and at any rate seems different from the one affecting main stress, where insertion is obligatory.

<sup>10</sup> Booij does not offer the syllabification of these words, but this can be inferred by statements regarding syllabification earlier in his book. For instance, Booij (1995: 36) reports that branching onsets cannot contain two sonorant consonants, banning combinations of nasals with liquids or glides, or liquids with glides. Given the consonant inventory of Dutch too (p. 7), /lj/ cannot be interpreted as the palatal lateral /ɭ/ either. For this reason, I take it that /lj/ genuinely refers to this consonant sequence, which, given the above, has to be syllabified as a coda-onset one. Marc van Oostendorp (p.c.), also confirms that this is the right syllabification.

stress is assigned from right-to-left forming a quantity-sensitive trochee, thus TROCHEE ('feet are trochaic') must be high-ranked. Secondary stress is assigned from left-to-right in a quantity-insensitive manner. Primary stress falls on the penult if it is heavy, i.e. CVV or CVC as in e.g. *a.xá:ta* 'Agatha' or *a.mán.da* 'Amanda'. Since CVCs are treated as heavy, MORAIC CODA (or WBYP) should be active, although with a caveat: it does not apply beyond the left of primary stress. WSP must therefore be quite high-ranking so that heavy syllables attract stress. The opposite also occurs, namely that stressed syllables are heavy suggesting that Stress-to-Weight (S-to-W) is also high-ranking.

If the penult is light, stress goes on the antepenult, as in *má:ra.tən* 'marathon'. Final stress is generally avoided, because NONFIN ('main stress is not word-final') dominates ALIGN-R (HDFT, PRWD) ('the head foot is at the right edge of the word'). Also NONFIN >> WSP holds, as the example (á:rən) rather than \*a(rón) indicates. There is an exception to final stress avoidance though: if the final syllable is superheavy, i.e. contains a CVVC or CVCC, then stress docks on that syllable e.g. *àt.mi.rá:l* 'admiral' or *lè:di.kánt* 'bed'. According to Gussenhoven, this can be accounted for by imposing the constraint Superheavy-to-Stress-Principle (SHSP), an extended version of WSP that refers to trimoraic syllables. If this is ranked above NONFIN, then we will correctly get final stress in *kaná:l* instead of penultimate \*kána:l. Finally, adjacent stresses are not allowed, therefore NOCLASH has to be dominant.

In the subsequent analysis, I will follow Gussenhoven (to appear) as faithfully as possible, although there are some weak - mostly technical - points in his analysis, such as the claim that WBYP does not work beyond the left of main stress or that FTBIN can be satisfied either at the moraic or syllabic level, allowing for instance a quadrimoraic bisyllabic foot such as (á:rən), which would by most analyses be considered two feet. All these however are tangential to the main point made here. Since I argue that ALIGN<sub>FO</sub> - my focus here - does not affect the stress algorithm, but merely turns onsetless syllables, which have been designated stressed, to onsetful ones, then the exact analysis of Dutch stress does not really matter, so long as it manages to produce the right stress. As far as I can see, what would most likely vary among different analyses is the footing assumed, but again this will pose no threat to the ensuing analysis. This is because ALIGN<sub>FO</sub> targets the stressed syllable independently of the footing. An analysis based on ALIGN<sub>FTO</sub> (Goedemans 1996, 1998) on the other hand might have a harder time in this respect.

### 3.3.2.2 The core analysis

It should be obvious that the data in (16) are compatible with the facts mentioned above. It is not the case that ALIGN<sub>σO</sub> affects the general stress algorithm. Had it done so, then since feet are formed from right-to-left and assuming ALIGN<sub>σO</sub> dominated WSP and NONFIN, we could perhaps expect to stress the word ‘aorta’ as *aɔrtá* instead of *aʔórta* since in the former the first onsetful syllable receives stress. This would be more or less what seems to happen in Juma mentioned in the discussion of the OSE effects, where the first onsetful syllable from the right receives stress. Dutch is not like that though. It will assign stress following the normal stress algorithm, but if it so happens that stress falls on an onsetless syllable in a hiatus environment, then it will insert a glottal stop. Thus, ALIGN<sub>σO</sub> must be dominated by the constraints mentioned before which are responsible for Dutch stress. This is illustrated below. For convenience, I group the constraints above under the tag: STRESS CONSTRAINTS.

(17) ?-insertion<sup>11</sup>

	/aorta/	STRESS CONSTR	ALIGN <sub>σO</sub>	DEP-?
a.	a.ɔr.(tá:)	*! (NON-FIN)		
b.	a.(ór.ta)		*!	
c.	a.(ʔór.ta)			*

The first candidate presents a case where ALIGN<sub>σO</sub> would be perfectly satisfied (as it happens in the winning candidate too), but would violate at least one of the constraints that regulate the stress algorithm in the language, e.g. NONFINALITY. The second candidate satisfies the stress constraints, but fails with respect to lower-ranked ALIGN<sub>σO</sub>. Because of the ranking ALIGN<sub>σO</sub> >> DEP-?, there is an additional candidate to be considered, namely one with glottal stop insertion, which actually proves to be the winner<sup>12</sup>.

This ranking can also neatly produce the absence of ?-epenthesis, when the onsetless syllable involved is stressless, as in *xáɔs*.

<sup>11</sup> I assume that the constraint MAX dominates at least DEP-? so that no input material is lost.

<sup>12</sup> Gussenhoven argues that feet in Dutch are either (HL) or (H). Examples (17b) and (17c) are footed as (HL) accordingly. The same holds for (20b) and (20c) below.

(18) lack of ?-insertion

/xaos/	STRESS CONSTR		ALIGN <sub>σ</sub> O	DEP-?
	NONFIN	WSP		
a. xa.(ós)	*!		*	
b. xa.(?ós)	*!			*
c. (xá:?.ós)		* (WSP)		*!
d. (xá:ós)		* (WSP)		

To understand why (18d) is chosen here, we need to explicitly make reference to the constraints NONFIN and WSP which are contained in the generic STRESS CONSTR group. Their relative ranking is NONFIN >> WSP as argued above. This comes in handy now. Candidates (18a) and (18b) stress the final syllable violating NONFIN which requires that final syllables are stressless. (18c) and (18d) avoid violation of NONFIN by stressing the penult. But this comes at a cost; since the final closed syllable is argued to be heavy, then it incurs a WSP violation by remaining unstressed. One way or another, (c) and (d) tie at this point, therefore the evaluation continues.

The crucial bit comes here; unlike (17), ALIGN<sub>σ</sub>O is vacuously satisfied in (18c) and (18d), because although they include an onsetless syllable, this is also unstressed. Since there is no requirement that stressless onsetless syllables acquire an onset, (18c) and (18d) escape violations of ALIGN<sub>σ</sub>O. As a result, the addition of a ?-onset as in (18c) gratuitously violates DEP-? and consequently loses. (18d) without ?-insertion is selected.

As it stands, this analysis has an undesirable result. It predicts that an initial onsetless syllable should acquire an onset if it is stressed. This is bound to happen, because of the ranking ALIGN<sub>σ</sub>O >> DEP-?, thus an example like [óli] ‘oil’ should actually surface as \*[?óli] instead. How can this problem be resolved? Positional faithfulness (Beckman 1999, Casali 1996) offers a way-out. Positional faithfulness has been designed to account for instances of positional neutralization and preservation of contrasts. It has been repeatedly found that a contrast may be retained or lost in all positions, but if it is to be lost in a specific environment, then this will be a weak position, and not a strong one. The general schema that produces contrast maintenance in a strong position only is given by:

(19) *Contrast in strong position only*

Faith<sub>Positional</sub> >> M >> Faith

This means that a contrast will be generally lost because Markedness dominates general Faithfulness, except for in a strong position, since Faithfulness relativized to this position protects the contrast from being lost. We can now transfer this schema to the case at hand substituting Faith with DEP-?, M with ALIGN<sub>σ</sub>O and Faith<sub>Positional</sub> with DEP-?<sub>INIT-σ</sub> (cf. McCarthy (2002: 32-36)). Empirically this means that in Dutch we should be able to find both onsetful and onsetless initially stressed syllables, but word-medially, stressed syllables should all be onsetful. The contrast between onsetful and onsetless stressed syllables is thus lost in the weak medial position, but maintained initially. The following tableau exemplifies.

(20) *No ?-insertion in stressed onsetless syllables word-initially*

	/oli/	STRESS CONSTR	DEP-? <sub>INIT-σ</sub>	ALIGN <sub>σ</sub> O	DEP-?
a.	o(lí:)	*! (NON-FIN)			
b.	(ʔó:li)		*!		*
c.	(ó:li)			*	

### 3.3.2.3 Remaining hiatus cases when $V_1$ is other than /a/

In the preceding discussion we had considered hiatus cases where the first vowel was /a/. We saw that if the second vowel was stressed, then ?-insertion would occur as a response to ALIGN<sub>σ</sub>O. If it was unstressed, then nothing would happen, since ALIGN<sub>σ</sub>O does not penalise onsetless unstressed syllables. The decision would thus be passed on to DEP-? which would militate against unnecessary epenthesis. In this section, we will examine what happens when the first vowel is other than /a/. As it will become evident from the data below, hiatus resolution take place, independently of whether the onsetless syllable is stressed or not. This is in contrast to the data before<sup>13</sup>.

(21) *Dutch glide-formation under hiatus when  $V_1 \neq /a/$  (Booij 1995: 66)*

a.	After front rounded vowels, glide is [ɥ]		
	/dyo/	[dýɥo]	‘duo’
	/edyard/	[édýɥart]	‘Edward’ (name)
	/røən/	[røɥən]	‘male dogs’
	/janyari/	[janyɥári]	‘January’
	/ryinə/	[ryɥínə]	‘ruin’

<sup>13</sup> I use the transcription of Booij, where /y/ is a front high rounded vowel and /j/ is the front unrounded glide. Dutch also possesses a front rounded glide /ɥ/ and a rounded labiodental back glide /ʋ/.

- b. After front unrounded vowels, glide is [j]
- |           |            |              |
|-----------|------------|--------------|
| /diet/    | [dijét]    | 'diet'       |
| /biɔskop/ | [bìjɔskóp] | 'cinema'     |
| /yeo/     | [yéjo]     | 'Geo' (name) |
| /zeən/    | [zéjən]    | 'seas'       |
- c. After back vowels other than /a/, glide is [ʊ]
- |          |           |               |
|----------|-----------|---------------|
| /ruanda/ | [ruuándə] | 'Rwanda'      |
| /boas/   | [bóʊas]   | 'Boaz'        |
| /hɔuən/  | [hóʊuən]  | 'hold' (verb) |

This distribution may at first glance seem odd compared to the previous facts. As we will see, it is not. The first 'peculiarity' is that no glottal stop is inserted. Instead, there is a glide, which also shares all its features with the first vowel, except that it is systematically [+high], even if the preceding vowel is [-high], as in e.g. *yejo*. Given that the majority of Dutch vowels prefer glide formation over ?-epenthesis, this suggests that glide formation is a more economical way to provide an onset than ?-epenthesis is. DEP-C must therefore dominate the constraint that gets violated during glide formation. This is INTEGRITY ('input segments do not have multiple correspondents'), because the input vowel has on the surface two correspondents, the vowel itself and the glide.

The second 'peculiarity' is that the creation of a glide onset here does not serve as a response to ALIGN<sub>MO</sub> satisfaction. Rather, an onset is provided independently of stress, thus ONSET >> INTEGRITY. But does this put the previous analysis in danger, i.e. does it predict we should get glide formation (or even ?-insertion) when the first vowel is /a/? I will show that it does not.

The idea is that ONSET will break the hiatus environment as much as it can by providing a glide onset due to ONSET >> INTEGRITY. However, as we have seen, this glide needs to be [+high], therefore onsets which are [-high] will not be admitted, meaning that \*[-high] ONS needs to dominate ONSET. This easily accounts for glide formation from high vowels. Since they are [+high], spreading all their features - including [+high] - will create the sought-after onset, only at the expense of low-ranked INTEGRITY, as in *dijet*.

But this now leaves us to consider what happens with the non-high vowels, i.e. mid and low ones. Of these, only mid vowels follow the glide formation pattern. The low ones do not. In fact, apart from adding a /ʔ/, the latter are also sensitive to stress placement. However, the glide formed after a mid vowel is not [-high], but predictably [+high], thus during the creation of the glide the [high] feature value changes. This is possible to do, if we assume that IDENT-[high], the constraint which requires that input

specifications with respect to the feature [high] do not change between input and output, is ranked very low (perhaps next to INTEGRITY). As a result, glide formation in the case of mid vowels will be admitted, because their [-high] feature will turn to [+high], which is characteristic of the glides, while their [-low] feature will stay intact. These effects are illustrated in the following tableaux.

(22)  $V_1 = \text{high vowel: glide formation}$

/diət/	*[-high] ONS	ONSET	INTEGRITY	IDENT-[high]
a. diét		*!		
b. dijét			*	

In (22), the first vowel is high, thus it can easily form an onset through glide formation. The winning candidate (22b) will thus only violate INTEGRITY.

(23)  $V_1 = \text{mid vowel: glide formation}$

/ze <sub>[-hi, -low]</sub> ən/	*[-high] ONS	ONSET	INTEGRITY	IDENT-[high]
a. zéən		*!		
b. zéC <sub>[-hi, -low]</sub> ən	*!		*	
c. zéj <sub>[+hi, -low]</sub> ən			*	*

Effectively, (23) results in the same output. The first candidate loses because it fails to present an onset that would resolve the hiatus. The other two candidates include a glide. The winner (23c) has a glide that is [+high], whereas (23b) presents a glide that is [-high]. While the latter is entirely faithful in terms of low-ranked IDENT-[high], it is already excluded by the high-ranking \*[-high] ONS, which militates against non-high onsets. Consequently, (23c) wins by only violating the low-ranked constraints.

Of course, one could apply the same strategy of changing the value of [-high] to [+high] in low vowels too. The problem is that the newly formed glide will be [+high, +low], i.e. an articulatorily impossible sound which can be easily eliminated by the high-ranked \*[+high, +low]<sup>14</sup>. But this is not yet sufficient to explain why no glide formation follows low vowels, since in principle not only the [high] feature could change its value, when the glide is formed, but also the [low] one, yielding a [+high, -low] glide after a low back vowel. This can be avoided if the constraint which preserves identity with respect to the feature [low], i.e. IDENT-[low] is highly-ranked.

<sup>14</sup> Other constraints which directly ban low glides are {A}=V that bans non-moraic counterparts of low vowels (Rosenthal 1994) or \*LOW GLIDE as in Smith (2005).



(24)  $V_1 = \text{low vowel}$ : no glide formation

	$/a_{[-hi, +low]} \text{orta}/$	$*[+high, +low]$	IDENT-[low]	ONSET	INTEGRITY	IDENT-[high]
a.	aʔorta					
b.	aorta			*!		
c.	aC <sub>[+hi, +low]</sub> orta	*!			*	*
d.	aC <sub>[+hi, -low]</sub> orta		*!		*	*

In (24), I have deliberately shaded the first candidate - the actual winner - as I would like to focus on the other candidates. More specifically, (24c) and (24d) are the candidates with some type of glide formation. Changing the feature value of [high] only, as in (24c), results in an unattested [+high, +low] configuration which is ruled out by the corresponding constraint. Changing the [low] value too (24d) avoids such violation, but incurs an equally fatal violation of IDENT-[low]. As it stands then, and without any additional modifications, the predicted winner should be (24b) which leaves the syllable onsetless. Some extra ingredient is obviously needed.

To sum up the discussion up to this point, what we have achieved so far is that glide formation occurs in a hiatus context irrespective of stress considerations, but only when the first vowel is high or mid. This is given by the ranking:  $*[+high, +low]$ , IDENT-[low]  $\gg$   $*[-high]$  ONS  $\gg$  ONSET  $\gg$  INTEGRITY, IDENT-[high]. This also tells us that when the first vowel is low, glide formation is impossible. But it does not yet tell us what happens instead. The findings of the previous section now become relevant. Recall that when the first vowel is low and the second vowel is stressed, ?-insertion in front of the onsetless syllable occurs. When that vowel is unstressed, nothing happens. The way we analysed that was by the ranking ALIGN<sub>σO</sub>  $\gg$  DEP-?, whereby glottal insertion was seen as a remedy triggered by ALIGN<sub>σO</sub> only. Since the addition of a glottal stop is allowed but only when there is no glide alternative, i.e. in the case of a sequence  $/a+V_2/$ , then  $*[+high, +low]$ , IDENT-[low] must certainly dominate DEP-?. However, the fact that ?-insertion only shows up in front of an onsetless *stressed* syllable and not in front of any syllable, indicates that ALIGN<sub>σO</sub>  $\gg$  DEP-?  $\gg$  ONSET. Now if we put all these together, what we get is:

(25) *Dutch ?-insertion and glide-formation under hiatus*

$*[+high, +low]$ , IDENT-[low], ALIGN<sub>σO</sub>  $\gg$  DEP-?,  $*[-high]$  ONS  $\gg$  ONSET  $\gg$  INTEGRITY, IDENT-[high]

What this means is that an onset needs to be created by means of a glide (ONSET >> INTEGRITY) provided the resulting glide is [+high, -low]. If it is something other than that, then glide formation is blocked due to \*[-high] ONS. However, this does not mean that the onsetless syllable will remain without an onset. An alternative strategy is now chosen, that of ?-epenthesis which is not normally preferred, since DEP-? >> INTEGRITY. While under DEP-? >> ONSET alone, one would not expect epenthesis at all, it is the force of the ranking ALIGN $\sigma$ O >> DEP-? that will enable epenthesis, but only under a certain condition, namely to provide an onset to an onsetless stressed syllable. Thus under the ranking in (25), we can express the fact that glide-formation occurs as a response to ONSET, whereas ?-epenthesis, as a response to ALIGN $\sigma$ O.

I will exemplify this full ranking with a few examples. I will start with the examples of section §3.3.2.1, and then move on to the cases discussed in the present section.

(26) When  $V_1 = /a/$  and (i)  $V_2 = \text{stressed}$ , (ii)  $V_2 = \text{unstressed}$

(i)	$/a_{[-hi, +low]} \sigma rta/$	*[+high, +low]	IDENT-[low]	ALIGN $\sigma$ O	DEP-?	*[-high] ONS	ONS	INTEGRITY	IDENT-[high]
☞ a.	a $\acute{\sigma}$ rta				*		*		
b.	a $\acute{\sigma}$ rta			*!			**		
c.	aC $_{[+hi, +low]} \acute{\sigma}$ rta	*!					*	*	*
d.	aC $_{[+hi, -low]} \acute{\sigma}$ rta		*!				*	*	*
(ii)	$/xa_{[-hi, +low]} \sigma s/$	*[+high, +low]	IDENT-[low]	ALIGN $\sigma$ O	DEP-?	*[-high] ONS	ONS	INTEGRITY	IDENT-[high]
☞ a.	xá: $\sigma$ s						*		
b.	xá: $\sigma$ s				*!				
c.	xá:C $_{[+hi, +low]} \sigma$ s	*!						*	*
d.	xá:C $_{[+hi, -low]} \sigma$ s		*!					*	*

Candidates (26c) and (26d) in both sub-tableaux are ruled out because they violate the top-ranked constraints. The difference then between (26i) and (26ii) lies in ALIGN $\sigma$ O. In the former it is active, therefore it chooses the candidate with ?-insertion as the winner (26i.a), but in the latter, it is vacuously satisfied by both remaining contenders, thus the decision is passed on to DEP-? which selects (26ii.a) as the winner.

(27) When  $V_1 \neq /a/$ : (i)  $V_1 = +high$  and  $V_2 = stressed (x)$  or  $V_2 = unstressed (y)$ ;

(ii)  $V_1 = -high$

(i.x)	/ryinə/	*[+high, +low]	IDENT-[low]	ALIGN 6O	DEP -?	*[-high] ONS	ONS	INTEGRITY	IDENT [high]
☞	a. ryɯ́ɪnə							*	
	b. ryʔɪnə				*!				
	c. ryɪnə			*!			*		
(i.y)	edyard	*[+high, +low]	IDENT-[low]	ALIGN 6O	DEP -?	*[-high] ONS	ONS	INTEGRITY	IDENT [high]
☞	a. édyɯart						*	*	
	b. édyʔart				*!		*		
	c. édyart						**!		
(ii)	/ye <sub>[-hi, -low]</sub> o/	*[+high, +low]	IDENT-[low]	ALIGN 6O	DEP -?	*[-high] ONS	ONS	INTEGRITY	IDENT [high]
☞	a. yéj <sub>[+hi, -low]</sub> o							*	*
	b. yéC <sub>[-hi, -low]</sub> o					*!		*	
	c. yéʔo				*!				
	d. yéo						*!		

The examples in (27) refer to cases where the first vowel is other than /a/. The common thing shared in all three sub-cases is that glide formation can avoid the violation of higher-ranked constraints such as \*[+high, +low] and IDENT-[low], since the vowels involved are other than /a/. As a result, the insertion of /ʔ/ now incurs a relatively high-ranking violation of DEP-? and proves fatal in all cases. Creating no onset at all, also violates ONSET, consequently the winning candidates are the ones that present glide formation<sup>15</sup>.

A final point that needs to be addressed is the directionality of glide formation. The data suggest that this is always rightward as shown below:

(28) *Rightward glide formation in Dutch* (Booij 1995, Rubach 2002)

a. /yeo/	[yejo]	*[yevo]	'name'
b. /poet/	[povet]	*[pojet]	'poet'
c. /dyo/	[dyɥo]	*[dyvo]	'duo'

In fact, leftward glide formation is not even employed as a remedy to avoiding glottal stop insertion, which as we have seen, is less preferred in the language. The example /kaunda/ → [kaʔúnda] rather than \*[kavúnda] illustrates this point. Rubach (2002) points out that there are languages which only allow rightward spreading (Dutch,

<sup>15</sup> The same result as in (27ii) would obtain, had the second vowel been stressed. The only difference would be that a candidate like (d) would be ruled out sooner, due to its violation of ALIGN 6O.

Malay), others which only allow leftward (Ukrainian) and others which allow both (Polish). As the specific analysis of this phenomenon is orthogonal to the larger picture examined in Dutch, I will follow Rubach and assume that the ranking ANTI-CRISP EDGE >> ONSET >> CRISP EDGE is in order. An informal definition of CRISP EDGE is that multiple linking between prosodic categories is prohibited (for a formal definition see Itô and Mester 1994). Thus, CRISP EDGE is violated when a segment spans across two syllables. ANTI-CRISP EDGE is treated as the reverse of CRISP EDGE. Due to the ranking mentioned above, satisfaction of ONSET will be preferred even if crisp edges are created, as the result of high-ranking ANTI-CRISP EDGE. This is what happens in rightward glide formation, i.e. /ia/ → [i.ja] (satisfaction of ANTI-CRISP EDGE), but not in leftward /ai/ → a.ji (satisfaction of CRISP EDGE). It should be noted that ANTI-CRISP EDGE, as the reverse of CRISP EDGE is stipulatory, thus while it serves for current purposes, a better solution seems plausible<sup>16</sup>.

### 3.4 Concluding remarks

In this section, I have examined the non-moraic effects between onsets and stress. Adapting Goedemans' ALIGN-FTO (1996, 1998), I have claimed that ALIGN $\phi$ O provides the appropriate account in generating both OSE and SOE effects. In the former, the presence of an onset attracts stress. ALIGN $\phi$ O affects prosodic aspects of the normal stress system of the language and causes misalignment of stress to an onsetful syllable. The ranking DEP-C >> ALIGN $\phi$ O >> ALIGN-STRESS produces this result. The opposite effect occurs when stress imposes the presence of an onset (OSE). Here, the stress algorithm remains unaffected, because it dominates ALIGN $\phi$ O, the only effect of which can be to add an onset to a syllable that has already been designated as the stressed one, in case it is onsetless. The ranking ALIGN-STRESS >> ALIGN $\phi$ O >> DEP-C captures this situation.

Before closing this chapter it is worthwhile mentioning that ALIGN $\phi$ O seems a counter-example to Blumenfeld's (2005c) hypothesis that markedness constraints relativized to prosodic domains will receive a segmental fix by removing the segmental markedness violation and not a prosodic fix by moving the boundary of the prosodic

---

<sup>16</sup> Rubach (2002) observes that CRISP EDGE and ANTI-CRISP EDGE might not be as unreasonable in the presence of other, apparently contradictory, constraints which have been used in the literature such as FINAL-C (McCarthy and Prince 1994) and FINAL-V (Rosenthal 1994). In addition to that, I would add that Alderete (1999) argues for the existence of anti-faithfulness constraints. While the (ANTI)-CRISP EDGE constraint is not of the faithfulness type, it is similar in spirit.

domain, unless the latter is covert, i.e. it involves no perceptual consequences at all. To visualize, he argues that if the avoidance of voiced obstruent codas is applicable, then in an example like /tab/, final devoicing [tap] will be preferred (segmental fix) rather than deletion [ta] (prosodic fix).

As we have seen previously, ALIGN<sub>σ</sub>O can be translated into a markedness constraint with reference to prosody in the shape of ONSET/σ (cf. Smith 2005) and become comparable to the type of constraints Blumenfeld uses, such as ASPIRATE/σ (“stressed syllables are aspirated”). If we now look into the strategies for ALIGN<sub>σ</sub>O/ONSET/σ resolution, it should be evident that both segmental and prosodic fixes are available. Onset insertion (Dutch) could under some interpretation be seen as a segmental fix, and thus should be valid. But if it is construed as a prosodic fix, like the deletion in /tab/→[ta] above, then it violates Blumenfeld’s claim. At any rate, the alternative solution, i.e. shift of prosodic domain, as it happens in the Aranda-type of languages, is unambiguously an allegedly impossible prosodic fix, given that it has perceptual consequences.

This concludes my discussion of onset-sensitive stress. From the next chapter onward (Ch. 4-7), I will shift my interest to other phenomena which are clearly weight-based and I will demonstrate that there too one finds effects from onsets. Since for such cases only an analysis that utilizes weight is possible, the proposal about the moraicity of onsets receives strong support.

## **Chapter 4**

# **Moraic onsets and Word Minimality in Bella Coola**

### **4.1 Introduction**

In this chapter I will focus on a clearly weight-based phenomenon, that of Word Minimality, and show that languages like Bella Coola calculate the onset for minimality purposes. As I have argued at length in chapter 1, for weight-based phenomena other than stress, no re-analysis along the lines of prominence is available, rendering cases such as Bella Coola Word Minimality as strong supporting evidence for the existence of moraic onsets.

Introducing weightful onsets in Bella Coola greatly improves Bagemihl's (1998) previous analysis of these data, which suffers from the undesirable designation of Word Minimality as an input condition and the necessary stipulation that all segments are underlyingly moraic. The current account is an output-oriented solution which solves these problems by admitting moraic onsets albeit in a very marginal environment. While onsets are generally non-moraic, as confirmed by the Root Maximality facts (cf. §4.2.2.1), they are moraic in /CV/ words only. In this way, the overarching requirement of bimoraic Word Minimality can be satisfied. Bella Coola thus presents a case of coerced onset weight, where onset moraicity is enforced by a higher imperative in the language, namely Word Minimality.

This chapter is structured as follows. In section 2 I introduce the data under investigation and summarize Bagemihl's (1998) analysis highlighting the problems it faces. In agreement with Bagemihl, I too argue that only an account based on moraic structure is feasible. In section 3 I show how this can be implemented avoiding Bagemihl's shortcomings. After outlining the core proposal, section 4 presents the analysis in detail and considers possible objections. Section 5 handles some residual issues and provides some tentative comments concerning the suitability of the MPARSE in accounting for the null candidates and the recognition of the morphological root as a prosodic domain. Section 6 offers a few concluding remarks.

## 4.2 Bella Coola basic facts

### 4.2.1 General syllabification and unsyllabified obstruents

In this section I will present some basic facts of Bella Coola as background for the Word Minimality analysis. Data and facts are based on Bagemihl (1991) and Bagemihl (1998) [henceforth B91 and B98 respectively]<sup>1</sup>. Note that much of this section will be devoted to reviewing the arguments Bagemihl (1991) has provided to justify the existence of unsyllabified obstruents that Bella Coola is so notorious for. This is important, because unsyllabified consonants play a key role in the overall analysis below. The Bella Coola inventory is presented below.

#### (1) *Bella Coola inventory*

[N.B: /c/=alveolar affricate, /ɬ/=lateral fricative, /ɬʷ/=lateral glottalic affricate]

p	t	c		k	k <sup>w</sup>	q	q <sup>w</sup>	
p'	t'	c'	ɬ'	k'	k' <sup>w</sup>	q'	q' <sup>w</sup>	ʔ
		s	ɬ	x	x <sup>w</sup>	χ	χ <sup>w</sup>	(h)
m	n		l	y	w			
m̥	n̥		l̥	i	u			
				a				

Despite allowing numerous unsyllabified consonants, as we will see in a moment, the language is quite usual in that it otherwise exhibits normal syllabification. Thus, syllables are maximally TRVVC, where *T*=*obstruent*, *R*=*sonorant* [I use *R* to represent a sonorant consonant which is either the second consonant in an onset cluster or is syllabic. This should be obvious from the context]. Only TR complex onsets are allowed, so TT and RR do not constitute complex onsets. Vowels and syllabic sonorants can serve as nuclei. Singleton obstruents syllabify either as onsets or codas depending on the environment. Codas receive moras, while onsets are claimed not to be moraic. The latter fact is generally true, but will be disputed later on, as we will see, in a particular instance. Some examples follow.

<sup>1</sup> Bagemihl uses a number of sources for his data. That is why I also mention the page where the example is located in Bagemihl's work so that the interested reader is able to check the original reference too.

(2) *Syllables with vocalic nuclei*

a.	CV	ʔ'i	"fast"	(B98: 75)
b.	CVC	k <sup>w</sup> it'	"to pry loose"	(B98: 75)
c.	CVVC	niiχ <sup>w</sup>	"fire"	(B91: 599)
d.	TRVC	xnas	"woman"	(B91: 601)
e.	TRVVC	c'wiiχ <sup>w</sup>	"having grey hair"	(B91: 619)

*Syllables with sonorant nuclei*

f.	CR	c'm̩	"index finger"	(B98: 79)
g.	CRC	tłq̩ <sup>w</sup>	"to swallow something"	(B98: 79)

As mentioned, obstruent clusters cannot syllabify as complex onsets, since TT complex onsets are banned. However, no deletion of consonants occurs either. Instead, in a string of obstruents - which can be as many as three or four root-internally and up to eight or nine across morphemes (cf. B98's examples on p. 74 and (55) here) - the one immediately before the nucleus syllabifies as an onset, the one immediately after a nucleus syllabifies as a coda, and the remaining consonants are unsyllabified<sup>2</sup>.

(3) *Sequences with unsyllabified consonants*

a.	st̩n̩	"tree"	(B91: 609)
b.	st'x <sup>w</sup> m̩	"floor mat"	(B91: 609)
c.	sq <sup>w</sup> cił	"ventral posterior fin"	(B91: 609)
d.	c'klakt	"ten"	(B98: 78)
e.	cipsx	"fisher"	(B98: 80)
f.	płtk̩n̩	"bark of bitter-cherry tree"	(B98: 79)

To establish the above finding, Bagemihl first shows that obstruents, contrary to sonorants and vowels, cannot serve as nuclei. He then supplies evidence corroborating the fact that in each syllable only one obstruent can syllabify as an onset and one as a coda. Any remaining obstruents must stay unsyllabified.

Data from reduplication, vowel allophony and glottal stop distribution are used to support the claim that sonorants and vowels pattern identically as nuclei to the exclusion of obstruents. For instance, roots with a vocalic or sonorant nucleus get reduplicated with one of the lexically-determined CV, CVC or V templates<sup>3</sup>, e.g. *qayt* → *qaqayt-i* "hat → toadstool-diminutive" with a vocalic nucleus (B91: 598) and *tłk<sup>w</sup>* →

<sup>2</sup> In this chapter, I will use **boldface** to represent unsyllabified consonants.

<sup>3</sup> Reduplication patterns in Bella Coola are complex, but as my focus is not reduplication, I will most of the time use examples of CV (the simplest) reduplication. Note, however, that reduplication facts may also prove relevant to the onset moraicity issue, but to keep things manageable I will leave this possibility open to future research.



*tltlk*<sup>w</sup> “swallow → continuative” with a sonorant one (B91: 599). If obstruents could be nuclei, then we would expect similar reduplication facts, e.g. *kt-* → *\*ktkt-* “fall” (B91: 606), but these are absent from the Bella Coola reduplication system. In fact, most *obstruent-only* words either do not reduplicate at all or they exhibit sonorant or vowel-epenthesis since the newly supplied nucleus allows them to get syllabified and reduplicated.

A number of arguments are also put forward against the syllabification of obstruent clusters. First, within syllables, a strict sonority profile is observed where sonority rises from the onset to the nucleus and then falls towards the coda. This is why complex onsets can only be TR - presenting rising sonority - and not TT or RR. TT clusters on the other hand are indifferent towards sonority considerations, cf. *sq<sup>w</sup>cił* (3c) or *płtkŋ* (3f), therefore they cannot be part of the syllable.

Reduplication facts are a very good source of additional arguments. TR clusters are syllabified in onset position. Had TT clusters been syllabified too, then we would expect them to pattern identically with TR clusters in processes such as reduplication. Words made of TT clusters only, as we have seen already, do not reduplicate at all or undergo nucleus epenthesis in order that they reduplicate (4a). On the other hand, TR clusters participate in reduplication as shown below (4b):

- (4) *Reduplication with TT-only sequences* (in the cases that this applies)<sup>4</sup>  
 a. *łq’-* (base) → *łŋq’* (n-insertion) → *ł<sub>n</sub>łŋq’-* (reduplication) “slap → continuative” (B91: 607)

*Reduplication in TR sequences*<sup>5</sup>

- b. *x<sup>w</sup>nał* → *x<sup>w</sup><sub>n</sub>x<sup>w</sup>nał-i*<sup>6</sup> “spring of water → diminutive” (B91: 615)

The really interesting facts though come into play when one considers what happens in the reduplication of words containing initial obstruent clusters followed by a nucleus. In

<sup>4</sup> Underlining is used to represent the reduplicant, while normal typeface is used for the base. Boldface is reserved for the unsyllabified consonants.

<sup>5</sup> It is obvious here that the reduplicant is copying the obstruent and the sonorant which now serves as the nucleus. This is merely a generalization though, as data are complicated when one considers other templates of reduplication (i.e. CVC or V patterns) as well as other processes that can accompany them, e.g. vowel syncope, glottal insertion. In the cases presented here, we might perhaps have expected a reduplicated form like (i) TRV.TRV or (ii) TV.TRV for a base /TRV/. The reason TR.TRV emerges as optimal could be attributed to Emergence of the Unmarked issues (McCarthy and Prince 1994), namely that although complex onsets are allowed generally in the language, reduplicants present the unmarked case banning complex onsets. This would exclude (i). (ii) would need to be excluded by a constraint that would explain why the original vocalic nucleus of the base is not used as the reduplicant’s nucleus too; the reason is that in that case, the sonorant would have to be skipped causing a Contiguity violation.

<sup>6</sup> Lengthening of the base vowel applies here [as in (6a) too]. This will not be of concern to us in the current work. For details, the interested reader is referred to Bagemihl (1991).

this case, reduplication is in fact possible, since there is already a nucleus available. However, when compared with reduplication of TR-initial sequences, a discrepancy arises. In TR clusters, the reduplicant appears prefixed immediately before the base TR sequence; and effectively, since the cluster is initial, the reduplicant appears in word-initial position. In TT clusters though, the reduplicant is not positioned in the same place, i.e. word-initially. Instead it is located before the last member of the cluster. Some examples are given below.

(5) *Reduplication in TR initial sequences*

- a.  $x^w na\dot{t} \rightarrow \underline{x^w n} x^w na\dot{t}-i$  “spring of water → diminutive” (B91: 615)

*Reduplication in TT initial sequences*

- b.  $p^{\prime} \dot{t} a- \rightarrow p^{\prime} \dot{t} a \dot{t} a$  “wink, bat the eyes → continuative” (B91: 609)  
 c.  $tq\dot{n} k- \rightarrow \underline{tq\dot{n} q} n k$  “be under → underwear” (B91: 609)

*Reduplication in a combination of clusters*

- d.  $skma \rightarrow \underline{skm} kma-y$  “moose → diminutive” (B91: 615)

The question then is whether there is any way to unify these patterns. In fact there is. It simply suffices to say that the reduplicant gets prefixed before the first base syllable<sup>7</sup>. In (5a) this coincides with the beginning of the word, since the TR clusters are part of the syllable. In (5b) and (5c), the reduplicant appears before the last obstruent (in the cluster) of the base, thus failing to align with left edge of the word. This comes as no surprise if we assume that the first consonant stays unsyllabified and remains outside the reduplication domain. (5d) illustrates the interaction of unsyllabified consonants and TR clusters. As it is now anticipated, the reduplicant consists of a copy of the onset obstruent and the sonorant that becomes syllabic. This is prefixed to the syllable and preceded by the unsyllabified consonant *s*.

Up to now, we have seen evidence against the onset syllabification of obstruents. No more than one obstruent can syllabify as an onset. How about codas? The same result obtains for codas too, as the reduplication facts again reveal.

(6) *Reduplication in final TT clusters*

- a.  $ya\dot{t} k- \rightarrow \underline{ya\dot{t} ya} a\dot{t} k-$  “do too much → continuative” (B91: 617)

<sup>7</sup> Bagemihl (1991: 613) based on examples such as *milix* → *milmilix* -*t**p* “bear berry → plant of the bear berry” (gloss mentioned in p. 603), argues that the reduplicant should be prefixed to a foot, rather than a syllable, since *mil* spans across two syllables. This is however not necessary in OT. All we need to say is that a CVC template needs to be filled from base segments without having to make specific reference to base syllabification. For present purposes at least, affixation to the first base syllable is sufficient.

- b.  $\chi^w alx \rightarrow \underline{\chi^w al} \chi^w alx$  “to melt  $\rightarrow$  solder” (B91: 617)

In these examples the postulated unsyllabified consonant is at the end of the word. If this could be incorporated in the preceding syllable, we would expect the reduplicated forms:  $*\underline{ya\dot{t}k}yaa\dot{t}k$  or  $*\underline{\chi^w al} \chi^w alx^8$ , but again the reduplicant prefixes before the syllable and fails to copy the final obstruent of the base, as this is not a part of the syllable<sup>9</sup>.

These data then confirm that the maximal syllable in Bella Coola is TRVVC. Only a single obstruent can syllabify in an onset and a coda position. Further obstruents are bound to stay unsyllabified.

One might however expect that these segments would get deleted, since they are neither syllabified nor extrasyllabic. Instead they surface. Bagemihl claims that although these segments are not licensed by syllables, they are nonetheless moraicly licensed and this suffices to allow them emerge. In other words, he claims that moras may license segments so that they receive a phonetic representation. As we will see in the next section, Bella Coola roots are subject to maximality restrictions which are computed on the basis of moras. Unsyllabified obstruents count for maximality purposes, a fact that confirms their moraicity.

As a final note here, bear in mind that since obstruents cannot act as nuclei, Bagemihl is also forced to claim that obstruent-only words will consist of no syllables at all, e.g.  $\dot{t}x^w$  “to wake somebody up”,  $c'kt$  “to arrive”,  $stx^w c'$  “cottonwood buds” (B98: 78). This statement is adopted here too. Unsyllabified obstruents and their moras will prove important in computations relating to Root Maximality and Word Minimality. These two phenomena are the topic of the next section.

## 4.2.2 Root Maximality and Word Minimality

### 4.2.2.1 Root Maximality

While our major focus with relation to moraic onsets is on Word Minimality, I will begin by discussing the Root Maximality facts, the characterization of which finds both

<sup>8</sup> Notice that in this instance, the reduplication pattern is a bimoraic one.

<sup>9</sup> Alternatively, one might suggest that the reason that forms like  $*\underline{ya\dot{t}k}yaa\dot{t}k$  are bad is because these complex codas are banned from reduplicants. In fact, we will see later on that the language generally lacks complex codas. This interpretation is indeed possible, but by no means provides a counter-argument to the explanation given in the text. The latter is consistent with the independently established fact that consonants remain unsyllabified under certain circumstances and thus offers a more unified approach.

Bagemihl (1998) and myself in agreement. These data will provide the basis of why mora-counting and not segment-counting is necessary to capture the Word Minimality data presented in §4.2.2.2.

As we have seen, Bella Coola has only a superficially complex syllable structure. In fact, maximal syllables are TRVVC with remaining consonants surfacing unsyllabified. What is elaborate though is the rich moraic structure that the language possesses. The major generalisation that can be drawn is that Bella Coola roots consist of maximally 4 moras [i.e. two bimoraic feet]<sup>10</sup>, where a root is a monomorphemic base to which affixes are added (B98: 91, note 8)<sup>11</sup>. Roots may also occur as independent words without any affixes (B98: 93, note 18). Bagemihl presents a large number of possible roots and corresponding examples that conform to such a generalisation. He also presents examples that would be considered to have five or more moras and these are all ill-formed (p. 77-81). To name a few<sup>12</sup>:

(7) *Heavy – Light syllable sequence* [B98: 79]

a. CVCCV	k'ucxi	“maggots”	3μ	[μμ][μ]
b. CVCCCV	λ'iq'łkŋ	“(low) dwarf blueberry”	4μ	[μμ]μ[μ]
c. CVCCCCV		*	5μ	[μμ]μμ[μ]
d. CVCCCCCV		*	6μ	[μμ]μμμ[μ]

Here the first syllable is closed, followed by an open syllable. Between the coda of the first syllable and the onset of the second one, only one unsyllabified consonant may intervene (7b). Additional ones would lead to an increase of moras beyond the total of four per root that the language accepts. Had the sequence been one of two closed syllables, then no unsyllabified consonants would be permitted, since the four-mora maximum would have been reached already as illustrated in (8).

(8) *Heavy – Heavy syllable sequence* [B98: 79]

a. CVCCVC	qacqit	“ant”	4μ	[μμ][μμ]
b. CVCCVC	*		5μ	[μμ]μ[μμ]
c. CVCCVC	*		6μ	[μμ]μμ[μμ]

<sup>10</sup> Yoruba (Ola 1995) is another language that posits a two feet maximum for prosodic words. Ola (1995: 282-3 and references cited therein) also mentions languages that make use of a two feet prosodic template. These include: Japanese, Ponapean, English and of course Bella Coola.

<sup>11</sup> Crucially, note that this is a root restriction and not a word one. Polymorphemic words frequently exceed this limit, e.g. *xłp'χ"łłłłłłs* “he had had in his possession a bunchberry plant” (B98: 74) [*xł-* “to have, possess”; *p'χ"ł* “bunchberry”; *-łł* “tree, plant”; *-łł* PLUPERFECT; *-s* POSSESSIVE). More examples are given in (55).

<sup>12</sup> Following Bagemihl’s notation, moras within brackets indicate syllabified moras, while those without any brackets flanking them are the moras contributed by unsyllabified segments.

A generalisation that becomes evident is that the more syllables a root has, the lesser space for clustering - and thus potentially unsyllabified consonants - is available. This can be shown when considering trisyllabic roots.

(9) *Trisyllabic roots* [B98: 81]

a. CVCVCV	quluci	“skunk”	3μ	[μ][μ][μ]
b. CVCVCVC	k’anawił	“bow of boat”	4μ	[μ][μ][μμ]
c. CVCCVCV	t’ixłala	“robin”	4μ	[μμ][μ][μ]
d. CVCCVCVC	*		5μ	[μμ][μ][μμ]

Here no unsyllabified consonants are permitted. In fact, no more than one heavy syllable can arise. Only one instance of a trisyllabic root with an unsyllabified consonant is imaginable and this presupposes that all syllables have to be light, so that there is space for one more moraic unsyllabified consonant, namely CCVCVCV, e.g. *stapiłm* “bat (animal)” (B98: 81).

Conversely, increasing clustering is possible in monosyllabic roots. These can maximally be CCCCVC when open, e.g. *płłkñ* “bark of bitter-cherry tree” (B98: 79), i.e. with three unsyllabified moraic consonants and one mora from the vowel, while CCCVC when closed, e.g. *q’χtis* “fish weir made of rocks” (B98: 79), or CVCCCC, e.g. *cipsx* “fisher” (B98: 80), with final clustering. Thus, extensive clustering is permitted either initially or finally, but not simultaneously at both edges.

(10) *Clustering at the edges in monosyllabic roots* (B98: 80)<sup>13</sup>

a. CCVCC	k <sup>w</sup> pałł	“liver”	4μ	μ[μμ]μ
b. CCCVCC	*		5μ	μμ[μμ]μ
c. CCVCCC	*		5μ	μ[μμ]μμ

Before concluding that the maximality criterion is found on mora count, we would need to dispense with any argument based on other types of counting, e.g. segment, consonant or vowel counting. For instance, to argue for segment counting, the four-mora maximum that Bagemihl persuasively presents, has to be interpreted in terms of segments, e.g. a root is maximally four segments long (or something longer than that,

<sup>13</sup> Bear in mind that while a form like CCCVCC (10b) is bad since it exceeds the mora maximum by virtue of the far too many unsyllabified obstruents, the near identical CCCVCC form is well-formed, e.g. *c’klakt* (11a). The difference here is that a complex onset that fits the sonority profile has been created. Since the cluster is syllabified in the onset, it no longer carries a mora, thus the total number of moras involved now reaches the maximum of four (i.e. μ[μμ]μ). This fact is an additional argument for the correct interpretation of the root maximality facts and their relationship to unsyllabified consonants and onsets.

but in any case always consistently so). This is impossible to do. Bagemihl (cf. B98: 83-84) points out that there exist roots with more than four segments, e.g. *stapit̪m* “bat (animal)” (B98: 78) or more than four consonants, e.g. *sq̣ʷχlun* “knee-cap”, so that a segment-based analysis cannot account for these deviations from an alleged four-segment maximum. In fact, no type of counting, i.e. segment-, C- or V-counting, apart from the moraic one can account for the facts in a uniform manner. This is illustrated below.

(11) *Possible counting criteria*<sup>14</sup>

	C-counting	V-counting <sup>15</sup>	Seg-counting	μ-counting
a. c'klakt	5	1	6	4
b. p'χʷt̪t	4	0	4	4
c. p̪t̪k̪n̪	4	1	5	4
d. λ'iq't̪k̪n̪	4	2	6	4
e. k'anawit̪	4	3	7	4
f. λ'aq̣ʷakila	4	4	8	4
g. miank	3	2	5	4

The well-formed roots above present variable numbers in terms of consonants, vowels or segments, but their moras are consistently four (the alleged maximum). However, given the table above the following maxima obtain for the rest of the categories: 5 consonants, 4 vowels or 8 segments.

It would be thus interesting to see whether these maxima generally produce well-formed roots as well. If any of them does, then we could claim that it is that criterion and not the mora one which is responsible for the observed root maximum. Assuming for instance that the language uses a segment-counting maximum (i.e. 8 segments), then we would expect roots with 8 segments to be consistently attested. The same should occur with the consonant maximum (5 consonants) or vowels (4 vowels).

<sup>14</sup> Glosses for words first time introduced here: *miank* “wide canoe” (B91: 616), *λ'aq̣ʷakila* “a man’s name” (B98: 78).

<sup>15</sup> I count syllabic sonorants as Vs by virtue of their function as nuclei. Perhaps nucleus-counting might seem more appropriate, but I am simplifying here and keep the V-counting in line with Bagemihl who considers sonorants as Vs, e.g. when discussing Word Minimality (cf. (13e)).

- (12) *Unattested roots based on C-, V- or Seg- maxima* [N.B: Boldface here is used to indicate that the maximum has been reached each time]

	C-counting	V-counting	Seg-counting	μ-counting
a. CVCCCCV (B98: 75)	<b>5</b>	2	<b>7</b>	<b>5</b>
b. CCCCC (B98: 75)	<b>5</b>	0	<b>5</b>	<b>5</b>
c. CCCVCVCV (B98: 81)	<b>5</b>	3	<b>8</b>	<b>5</b>
d. CCVCCVC (B98: 80)	<b>5</b>	2	<b>7</b>	<b>5</b>

(12) however shows that many roots that conform to these requirements are nevertheless unattested. All the roots in (12), have 5 consonants - which in (11a) is acceptable - and are unattested. So is the case in (12c) that has 3 vowels, although in (11e), this does not entail any problems. Under the segment counting approach, 8 segments do not ensure the well-formedness of the root, as (12a)-(12d) illustrate. In fact (12b) has 5 segments only, but still does not occur. Thus, even though these examples contain a number of units which other extant roots exhibit, they do not appear as well-formed roots. One thing unites them however; they all exceed four moras. Therefore, the root maximum can only be defined in terms of moras.

Having seen that an analysis based on a different type of counting other than mora-counting is unsuccessful, let us consider a final alternative, namely that of an extraprosodicity account. Bagemihl correctly observes that such a solution would also be untenable, because it predicts that unsyllabified consonants should be able to arise freely at both word-edges. The problem arises once we consider the following. If there are Bella Coola roots which contain up to three such consonants at the left edge, e.g. *pʰtkɲ* “bark of bitter-cherry tree”, and up to two at the right edge, e.g. *ʔastx* “(to be) inside” (B98: 84), then why should these not appear simultaneously forming e.g. \*CCC.CVC.CC? This seems coincidental unless one considers that structures like the above exceed the four-mora maximum.

In conclusion, the Root Maximality facts demonstrate that nuclei, codas and unsyllabified consonants carry moras, whereas onsets do not. The overwhelming majority of Bella Coola roots conform to the 4-mora limit (over 94% out of the 1169 monomorphemic roots listed in Nater (1977)). Since the exceptions are either personal or geographic names, loanwords or possibly morphologically complex forms, the generalisation put forward seems well-grounded (see Bagemihl 1998, Appendix B for the exceptions).

#### 4.2.2.2 Word Minimality

In addition to the maximality effects, the language also exhibits a word minimum effect. As we will see later on, it is there that Bagemihl's analysis no longer flows as convincingly. The minimal word facts are stated below (B98: 87-89)<sup>16</sup>:

(13)	Word Shape	Examples
a.	V	*
b.	VV	ya <sup>17</sup> "good"
c.	VC	ɲʌ' "dark, night"
d.	C	*
e.	CV / CR	ʌ'i "fast" / c'm "index finger" (B98: 79)
f.	CC	tk' "sticky"
g.	CCC	sχp "to tie a knot"
h.	CCCC	p'χ <sup>w</sup> ʌt "bunchberries"
i.	CCCCC	*

Before we proceed, we need to clarify that by CC clusters, Bagemihl refers to obstruent clusters, and not for instance to a sequence of an obstruent followed by a sonorant which can act as a nucleus. The latter case falls under what he considers CV sequences (13e). This means that the consonants in (d) and (f)-(i) remain unsyllabified, since none can act as a nucleus.

Bagemihl - as we do as well - claims that the word minimum is based on the notion of mora. Examples (13a) and (13d) are unattested since they are monomoraic, whereas (13i) cannot surface as it exceeds the four-mora limit that Bella Coola roots need to observe. The remaining cases are all at least bimoraic and thus correctly surface.

Suppose however we contest Bagemihl's claim that final long vowels are rare and let us consider a genuinely long vowel, e.g. *aa* and not *ya* of (13b). Given that this consists of one root node, it can be treated as a single segment alongside words consisting of only V or C. With this in mind, it could be argued that the Minimal Word

<sup>16</sup> Notice that since roots are the only morphemes which can occur independently as words without any affixes, a minimal constraint on roots is equivalent to a minimal-word constraint (Bagemihl 1998: 93).

<sup>17</sup> Bagemihl (1998: 89) presents this as an example of VV and later on (1998: 93) mentions that there are no examples of roots consisting of only a long vowel, which may be an accidental gap, but such absence is more likely attributed to the extreme rarity of long vowels in morpheme-final position (Nater 1984: 17). Although a bimoraic structure for *ya* is a possible one [cf. Smith (2003: example (8b.ii))], one might still argue that such an example constitutes a CV sequence and not a VV one. In this case and under a moraic approach, the lack of VV words is unexpected. Nonetheless, the extreme rarity of morpheme-final long vowels mentioned above could be employed for such absence. Note though that in the segment-based approach below no similar idea has to be postulated. There, VV can be treated as a single segment, if one segment means one root node. This is however a small gain, unable to compensate for the serious drawbacks the segment-based approach carries along.



in Bella Coola merely has to consist of two segments. Since all of V, C and VV include just one segment, it would be correctly predicted that these fail the minimality criterion, and as such never surface as possible roots. In the light of the above, could we then successfully pursue the claim that the language simply requires minimally bisegmental words? The answer will be negative.

First, this proposal is undesirable since no languages seem to be sensitive to such segmental restrictions when considering word minimality<sup>18</sup>. It would thus be plausible to be able to frame the problem within a more cross-linguistically likely solution, such as the one advanced by the independently established and well-accepted moraic theory. As we will see in a moment, this is indeed possible, as we only need to posit a bimoraic word-minimum for Bella Coola. So in this respect, the language is one among many others that impose such a requirement and does not present an exceptional or unusual case.

There is additional supporting evidence against a segment-based analysis particular to Bella Coola. For instance, as noted before, the language exhibits rich reduplication in the form of CV, V and CVC reduplicants (Bagemihl 1991). It suffices to mention here that these can plainly be described as consisting of either one or two moras, whereas no adequate generalization can be achieved in a segment counting approach.

Moreover, apart from the Word Minimality restriction, we have already seen (§4.2.2.1) that the language also has a Root Maximality constraint which states that roots consist of maximally four moras. If a segment-based analysis is put forward for the minimality case, one would prefer to maintain a uniform analysis for the maximality

---

<sup>18</sup> One potential exception I am aware of is Yakima Sahaptin (Curtis 2003: 193-207), where minimal words seem to be of the type CCV or CVC. CV words do not occur, but surprisingly CVV syllables although admitted in the language do not satisfy the minimality criterion either. Curtis suggests that if one insists to see this as a prosodic phenomenon, then the minimality criterion must be at least two consonants. Given that this is unsatisfactory under standard weight approaches, she suggests that minimality is not a prosodic phenomenon, but instead “is related to lexical recoverability and the overwhelming frequency of consonant clusters in Sahaptin” (2003: 206). I find this equally unsatisfactory but also rather cryptic and unclear, so I opt for a tentative suggestion made to me by Moira Yip (p.c). Given that all initial clusters are invariably released (Curtis 2003: 196), some illicit clusters are split by epenthetic or excrescent [i] (Curtis 2003: 195), and that some codas are released (Curtis 2003: 198, 199), it makes sense to suggest that perhaps CCV and CVC sequences are more accurately something like C<sup>i</sup>CV and CVC<sup>i</sup> respectively. Under this conception, the minimality criterion can be interpreted as disyllabic. It will not come as a surprise then that CVV words are missing, since they are made up of just one syllable. Hence, Sahaptin is not a counter-example to the prosodicity of the minimal word phenomenon. What’s more, it presents itself as an additional language - apart from Bella Coola - where entities usually thought of as unable to contribute to weight, actually may do so, as is the case with onsets in Bella Coola or excrescent vowels in Sahaptin.

effects too<sup>19</sup>. But as we have seen in the preceding section, only a mora-based account can explain the root maximality facts, thus the only option is to propose a uniformly moraic approach.

#### 4.2.2.3 Problems for Word Minimality in Bagemihl (1998) and a solution

Under the view of a moraic analysis, a problem now becomes evident when one compares V words which do not meet the WdMin requirement with the CV ones which do. Given Bagemihl's analysis, both should be light monomoraic by virtue of the single mora contributed by the vowel. However, CV satisfies the word minimum, while V does not. To resolve the problem, Bagemihl assumes that all segments are underlyingly moraic so that V syllables are underlyingly monomoraic, while CV ones are underlyingly bimoraic. Similarly, C is monomoraic and CC is bimoraic. If WdMin is allowed to hold in the *input*, then it is correctly predicted that CV and CC satisfy it, while V and C do not. In the output however, CV and V are monomoraic - since the singleton C syllabifies as an onset, it cannot retain its mora - and syllabified, while C and CC share the fact that remain unsyllabified (although the former is monomoraic and the latter bimoraic). The representations Bagemihl proposes are (1998: 88-89)<sup>20</sup>:

(14)	Word Shape	Respects WdMin?	Prosodic Structure	
			Underlying	After syllabification
a.	C	No	$\mu$	$\mu$
b.	CC	Yes	$\begin{array}{c} \mu \quad \mu \\   \quad   \\ C \quad C \end{array}$	$\begin{array}{c} \mu \quad \mu \\   \quad   \\ C \quad C \end{array}$
c.	V	No	$\mu$	$\begin{array}{c} \sigma \\   \\ \mu \end{array}$
d.	CV	Yes	$\begin{array}{c} \mu \quad \mu \\   \quad   \\ C \quad V \end{array}$	$\begin{array}{c} \sigma \\   \\ \mu \\ / \quad \backslash \\ C \quad V \end{array}$

<sup>19</sup> Alternatively, one could utilize a segment-based approach for WdMin facts and a mora-based one for RtMax effects. A segment-based approach adds an extra component to the analysis and implies that cross-linguistically there can be other languages that base their WdMin considerations on segment counting. This seems unsupported. It is thus preferred to use a moraic approach throughout achieving consistency and making a less controversial claim, i.e. that onsets may be moraic, a fact suggested by data in several other languages and phenomena examined in this thesis.

<sup>20</sup> Recall that (14a) and (14b) include unsyllabified segments and thus come without any syllable node. They manage to surface though because they are moraicly licensed (see end of §4.2.1).

These representations bring forward two issues, the latter being more serious; first, (14a) and (14b) do not respect the Strict Layer Hypothesis (SLH), since no syllable dominates (any of) the consonants. This may be seen as a significant drawback, because the SLH seems to be universally applicable, at least in the sense that for each word, there is at least one foot, which dominates at least one syllable, which dominates at least one mora. Apparently though, Bella Coola presents direct evidence that the SLH can nevertheless be violated, if it conflicts with higher demands of the language, namely what can constitute an acceptable nucleus in the language.

More importantly, for current purposes, Word Minimality has to be treated as a condition in the *input*, because (14d) satisfies it and (14c) does not, and yet in the output they each have exactly one mora. This cannot be maintained in an output-based theory, where such restrictions spring from constraint interaction. The analysis to be presented next attempts to resolve this issue by claiming that the reason (14d) respects Word Minimality and (14c) does not, is because onsets in Bella Coola are allowed to be moraic in the output, albeit in a very restricted environment. If this possibility is admitted, then the C of CV will be treated as a moraic onset, hence (d) will actually be bimoraic in the *output*. Word Minimality will no longer need to be hypothesized as a property of the input. As for CC, this manages to satisfy word minimality, despite being unsyllabified. Constraints regulating the minimal sonority of the nucleus, e.g. \*Nuc/Obstr >> Nuc >> \*Nuc/Son >> \*Nuc/Vowel, will ensure that obstruent-only words will remain unsyllabified. Word Minimality though is only interested in the moraic make-up of the words; hence bimoraic obstruent-only words will still satisfy it. The modified representation proposed is depicted below<sup>21</sup>:

---

<sup>21</sup> Unlike Bagemihl, I am being neutral as to whether the consonants should have any moras underlyingly (more on that in the beginning of section §4.3). This is why consonantal moras are within brackets.

(15)	Word Shape	Respects WdMin?	Prosodic Structure	
			Underlying	After Syllabification
a.	C	No	(μ)	μ
b.	CC	Yes	(μ) (μ)     C C	μ μ     C C
c.	V	No	μ	σ   μ   V σ
d.	CV	Yes	(μ) μ     C V	σ / \ μ μ     C V

As mentioned, onsets will be moraic *only* in this extremely restricted environment, namely *CV words*. We can tell that in larger roots, this is not the case due to the maximality restrictions discussed in §4.2.2.1. In particular, roots such as CCVCC *sqʷatkʷ* “ashes” [B98: 76] or CVCCC *muχʷt* “to cry huyp (a dance cry)” [B98: 75] or CVCCCV *nuχʷski* “soapberries” [B98: 75] among others are admitted, because all of them are maximally quadrimoraic. Had the onsets counted moraically, all these forms would either include 5 or 6 moras. Given the maximality restrictions, they should thus fail to surface. The fact that they surface, clearly suggests that onsets are not moraic with the exception of CV words<sup>22</sup>. The following sections present a detailed analysis of the above.

### 4.3 The core analysis

The current proposal focuses on the word minimality facts and then expands on the full range of data to provide a complete analysis. It maintains much of Bagemihl’s original analysis, but admits moraic onsets in just one particular environment. Note though that unlike Bagemihl who assumes that moras are underlyingly present after Hyman (1985),

<sup>22</sup> Alternatively, one may propose that onsets are consistently moraic and the only thing that needs to be done is to revise the mora maximum to five or six moras so that moraic onsets are also taken into account. However, this is not correct. To illustrate, suppose we modified the limit to six moras, so that roots like CVCCCV above are admitted [i.e. with all segments moraic]. The question would then be why roots such as CCCCCV or CVCCCC or CCVCCC [B98: 79-80] etc. are not well-formed, although these comprise six moras too. Such an amendment would ruin the generalisation formerly established, which as we have seen is well-grounded. Evidently then onsets are moraic only in CV roots and the moraic limit should not be modified.

I will present an analysis which works irrespective of the presence of underlying moras, so as to conform to the Richness of the Base (Prince and Smolensky 1993/2004).

However, because the language has phonemic vowel length and vowels have to be distinguished in this respect, I will represent short vowels as monomoraic and long ones as bimoraic. This is an assumption that every analysis needs to make. More generally, even Rosenthal (1994) who argues that vowels acquire moras through the constraint V-mora<sup>23</sup>, claims that long vowels are lexically specified in terms of two moras and are not subject to V-mora.

Unsyllabified consonants will also be admitted, while the only complex onsets allowed are of the obstruent-sonorant kind. Presumably a ranking like SonSeq >> \*Complex Onset holds, but this will not be discussed here in more detail.

The core of the analysis deals with the environment where moraic onsets appear; it argues that moraic onsets only emerge so that Word Minimality, a top priority in the language, is satisfied. This renders CV words minimal. Larger forms already satisfy Word Minimality, thus use of moraic onsets is superfluous. Moreover, as the language also presents a Root Maximality condition, we are actually able to tell that onsets do not contribute any moras in other cases. If they did, then we would predict that certain attested roots like CVC.CVC should instead be unattested, because they would exceed the four-mora maximum.

In due course, numerous technical details will be added to complete the analysis. While these details might be more relevant to the OT theorist, careful technical development is crucial for empirical reasons too. For instance, I will devote detailed discussion on why the output of a /CV/ word appears moraified and syllabified as  $[[C_\mu V_\mu]_\sigma]_{w_d}$ . This is because there is independent evidence from reduplication that CV forms get syllabified (cf. §4.2.1). It is therefore important that all other reasonably conceivable candidates such as:  $[C[V_\mu]_\sigma]_{w_d}$ , the null parse,  $[C_\mu V_\mu]_{w_d}$  and  $[C_\mu[V_\mu]_\sigma]_{w_d}$  examined in (19), (28), (36), and (38) respectively are eliminated. Similarly, we also know that words that consist of a single C or V do not occur either. Therefore, in this case too, we need to dispense with alternative possibilities (see also some relevant discussion in fn. 36 and §4.4.4). Both these requirements reflect empirical facts that require an elaborate analysis, which will be developed shortly. For the time being however, I will only lay out the basic proposal.

---

<sup>23</sup> The V-mora constraint states that for every vocalic root node  $rt_i$ , there is a mora  $\mu_i$  (Rosenthal 1994: 26 in ROA-126 version). Effectively, this says that every vowel is (mono-)moraiac.

The form we need to examine is /CV/. Unlike Bagemihl who needs to make two crucial assumptions, namely that: i) Word Minimality is an input condition and ii) all segments are underlyingly moraic, in the current proposal, segments do not have to be underlyingly moraic and Word Minimality is not an input condition but an output constraint.

(16) WDMIN: Words are minimally bimoraic

Given that WDMIN is a requirement that is never violated in the language, it is reasonable to assume that it is undominated. The proposed solution admits moraic onsets, therefore the constraint that militates against them (17), has to be dominated by WDMIN.

(17) \*MORAIC ONSET: Moraic onsets are banned

Since we want to be able to consider both moraically specified and unspecified inputs, i.e. /C<sub>μ</sub>V<sub>μ</sub>/ and /CV<sub>μ</sub>/ respectively<sup>24</sup>, we want to permit mora insertion as a strategy for WDMIN satisfaction. As a result, DEP-μ must be dominated by WDMIN too. We therefore obtain the following ranking:


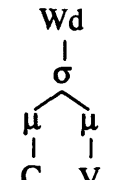
(18) WDMIN >> DEP-μ, \*MORAIC ONSET

Let us now see how /CV/ fares in terms of these constraints. Importantly, note that here and throughout, the brackets surrounding the mora indicate that if they are considered, then the input is assumed to include a mora. In the case of a non moraic input, violations of DEP-μ become relevant. These are also represented within brackets. The remaining violations are shared in both moraic and non-moraic inputs. This practice will be repeated again whenever it does not overburden the tableaux.

---

<sup>24</sup> Since vowels are moraically specified either as monomoraic (short) or bimoraic (long), the terms moraicity and non-moraicity refer to consonants.

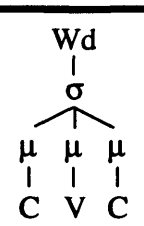
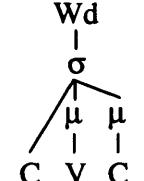
(19) /C<sub>(μ)</sub>V<sub>μ</sub>/ --- [[C<sub>μ</sub>V<sub>μ</sub>]<sub>σ</sub>]<sub>PrWd</sub>: WDMIN >> DEP-μ, \*MORAIC ONSET

	WDMIN	DEP-μ	*MORAIC ONSET
a. 	*!		
b. 		(*)	*

An output like (19a) violates WDMIN. The candidate in (19b) has a moraic onset, but this violation is less important than the one incurred by the first candidate. An additional DEP-μ violation occurs when the input is /CV<sub>μ</sub>/. The outcome remains unaffected since WDMIN >> DEP-μ.

In a form larger than the CV one, WDMIN can be satisfied by other means, so our claim is that no moraic onsets will appear in such a case, because although low-ranked, \*MORAIC ONSET violations are taken into account. Even if the input is moraically specified, \*MORAIC ONSET dominates MAX-μ, the constraint responsible for mora preservation. As a result, moras will not appear on onsets even at the expense of mora loss. All these can be illustrated in the following example<sup>25</sup>.

(20) /C<sub>μ</sub>V<sub>μ</sub>C<sub>μ</sub>/ --- [[CV<sub>μ</sub>C<sub>μ</sub>]<sub>σ</sub>]<sub>PrWd</sub>: WDMIN >> DEP-μ, \*MORAIC ONSET >> MAX-μ

	WDMIN	DEP-μ	*MORAIC ONSET	MAX-μ
a. 			*!	
b. 				*

<sup>25</sup> For full ranking of /CVC/ input (moraic and non-moraic), see Appendix A example (20).

To illustrate this point in (20), I consider the fully moraically specified candidate  $/C_{\mu}V_{\mu}C_{\mu}/$ . Its  $/CV_{\mu}C/$  counterpart would just present some DEP- $\mu$  violations without changing anything in the outcome. Both candidates satisfy WDMIN, but only the first one includes a moraic onset. Due to the proposed ranking, (20a) incurs more severe constraint violations than (20b) and hence loses. Of course, empirically (20a) cannot be distinguished from (20b). However, Root Maximality facts support a representation like (20b).

Recall from §4.2.2.1 that roots in Bella Coola consist of maximally four moras. This is encoded in the following undominated constraint.

- (21) **RTMAX**: No root may exceed four moras [B98: 77] (for an amendment see beginning of section §4.4.1)

RTMAX sets a limit to the number of moras, so if this is superseded, no fully-fledged output may win. Instead, the null parse is selected.

- (22) **MPARSE**: Morphemes are parsed into morphological constituents [Prince and Smolensky 1993/2004]

This is briefly illustrated here (various details are added in section §4.4.1).

- (23)  $/C_{(\mu)}C_{(\mu)}C_{(\mu)}V_{(\mu)}C_{(\mu)}C_{(\mu)}/$  ---  $\emptyset$ : RTMAX >> MPARSE

	RTMAX	MPARSE
a. <p>The diagram shows a root node 'Wd' branching into three nodes: a syllable node 'σ', and two moras 'μ'. The syllable node 'σ' branches into a moraic onset 'μ' and a moraic nucleus 'μ'. The moraic onset 'μ' branches into three consonants 'C C C'. The moraic nucleus 'μ' branches into two consonants 'C C'. The two moras 'μ' from the root node 'Wd' branch into two consonants 'C C'.</p>	*!	
b.		*

Let us abstract away from the exact input mora specification and focus on the two candidates presented here. Anticipating the relevant discussion, let us also assume that unsyllabified consonants only survive if they bear a mora. Then, as (23a) shows, a  $/CCCVC/$  root fails to emerge. This is because it includes three unsyllabified consonants, all of which need to be moraically licensed in order that they survive. But



this means that the root maximum is exceeded, since the CVC syllable is already bimoraic. Candidate (a) therefore is ruled out and the null parse is considered optimal.

Notice however that a root such as CCVCC (24) with just one less unsyllabified consonant is well-formed, e.g. *kʷpaH* “liver” (B98: 80). What this means is that this root adheres to RTMAX and implies that the onset of the CVC syllable has to be non-moraic.

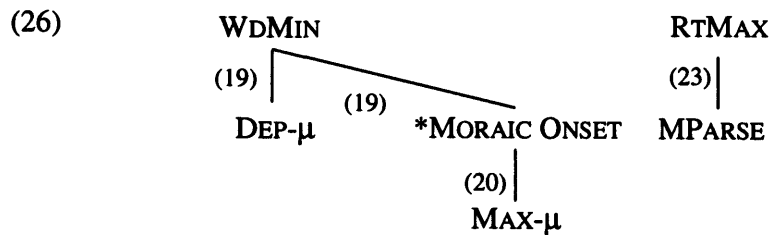
(24)  $/C_{\mu}C_{\mu}V_{\mu}C_{\mu}C_{\mu}/$  ---  $[C_{\mu}[CV_{\mu}C_{\mu}]_{\sigma}C_{\mu}]_{PrWd}$ : RTMAX, \*MORAIC ONSET >> MPARSE

	RTMAX	*MORAIC ONSET	MPARSE
a. <div style="text-align: center;"> <math>W_d</math>  <math>\swarrow \quad \downarrow \quad \searrow</math>  <math>\sigma</math>  <math>\swarrow \quad \downarrow \quad \searrow</math>  <math>\mu \quad \mu \quad \mu \quad \mu \quad \mu</math>  <math>  \quad   \quad   \quad   \quad  </math>  <math>C \quad C \quad V \quad C \quad C</math> </div>	*!	*	
b. $\emptyset$			*!
c. <div style="text-align: center;"> <math>W_d</math>  <math>\swarrow \quad \downarrow \quad \searrow</math>  <math>\sigma</math>  <math>\swarrow \quad \downarrow \quad \searrow</math>  <math>\mu \quad \mu \quad \mu \quad \mu</math>  <math>  \quad   \quad   \quad  </math>  <math>C \quad C \quad V \quad C \quad C</math> </div>			

Note that no (direct) ranking argument can be formed between RTMAX and \*MORAIC ONSET, since the winning candidate violates both. Later on, however, we will be able to indirectly infer that RTMAX >> \*MORAIC ONSET, through the interaction of other constraints. At this stage, the inclusion of this tableau merely highlights that apart from reasons specific to our analysis suggesting that supraminimal words contain no moraic onsets (cf. (20)), there is also empirical evidence leading to this conclusion. Roots of the type in (24) are acceptable, so clearly the upper limit of moras per root entails that moraic onsets do not appear in maximal roots.

Summarising the core part of this analysis, we have provided the ranking arguments given in (26) and have proposed a mini-grammar for Bella Coola which is consistent with the one in (25).

(25) RTMAX, WDMIN >> MPARSE >> DEP- $\mu$ , \*MORAIC ONSET >> MAX- $\mu$



The exact ranking of these constraints as indicated in (25) will be justified as the analysis progresses. Additional constraints will also be employed, but these will just be needed to polish the analysis so that various details are covered. The next section ranges over Word Minimality and Root Maximality facts and examines them in detail.

## 4.4 Detailed analysis

### 4.4.1 Word Minimality

I will begin by examining the word minimality facts. I assume that the foot node is always there, therefore its absence in the following representations should not be construed a violation of SLH (for more on this see (34)).

In fact, there is supporting evidence for the presence of the foot node. According to Kiparsky (2002), “moras which are not affiliated with syllables or feet do not count towards syllable weight or foot size”. Nonetheless, moras of unsyllabified Bella Coola consonants do contribute to the computation of the root maximum of four moras. If Kiparsky is right, then these moras have to link directly to the foot level. An amendment of RTMAX is thus in order. The constraint should require that *roots maximally contain two feet*. However, due to the absence of footing evidence, e.g. stress, we do not know how exactly the moras should be footed. Thus for expository (and space) reasons, I simplify the analysis by disregarding feet and continuing to compute RTMAX violations in terms of moras. Also bear in mind that the root maximum is exactly four moras (or exactly two feet), so no additional moras can be left unaffiliated or adjoined to higher prosodic structure.

Let us first consider the case of monosegmental words. These fail to emerge (cf. (15a) and (15c)). Since they are subminimal forms, they violate WDMIN. However, unlike /CV/ forms which circumvent the same problem by assigning a mora to their onset, /C/ or /V/ forms are not rectified in any analogous way. An obvious remedy for a monoconsonantal form would be the insertion of a vowel, but this is not allowed,

therefore the null parse is preferred. This entails that DEP-SEG - which penalises segment insertion - is highly-ranked and dominates MPARSE.

(27) /C<sub>(μ)</sub>/ --- [Ø]: WDMIN, DEP-SEG >> MPARSE<sup>26</sup>

	WDMIN	DEP-SEG	MPARSE
a. $\begin{array}{c} \text{Wd} \\   \\ \sigma \\   \\ \mu \\   \\ \text{C} \end{array}$	*!		
b. $\begin{array}{c} \text{Wd} \\   \\ \sigma \\ / \quad \backslash \\ \mu \quad \mu \\   \quad   \\ \text{C} \quad \text{V} \end{array}$		*!	
c. $\emptyset$			*

Things are slightly more complicated when one considers monovocalic forms. Here, apart from the insertion of a segment, e.g. a consonantal onset, which would already be excluded due to DEP-SEG, another option is feasible, namely lengthening of the vowel so that it becomes bimoraic. In fact, this idea puts the analysis proposed so far at risk. The reason is that if we are to avoid vowel lengthening, we could employ violations of DEP-μ. However, we have already seen that DEP-μ (cf. (19)) can be violated in Bella Coola, so it must not be too highly ranked. If the solution lies in the use of DEP-μ then the only thing we can suggest is that DEP-μ >> MPARSE so that the null parse makes a better candidate than the one with the lengthened vowel.

The general facts though run against such a hypothesis, because otherwise we would expect the null parse to be the winner more often than it is. For instance, for the input /CV<sub>μ</sub>/ which we have already encountered, DEP-μ >> MPARSE predicts the null parse as the winner, but since we know that CV words are attested, it must be the case that MPARSE >> DEP-μ.

<sup>26</sup> Although I will be arguing that the null parse is the winning candidate for /V/ and /C/ roots as well as for roots that exceed the mora maximum imposed in the language, other logical possibilities are also available, e.g. /V/ could lengthen to [VV] and /C/ could become [CV] by epenthesis. This issue is addressed in more detail in fn. 36 below and in §4.4.4.

(28) /CV<sub>μ</sub>/ --- [[C<sub>μ</sub>V<sub>μ</sub>]<sub>σ</sub>]<sub>PrWd</sub>: MPARSE >> DEP-μ

	MPARSE	DEP-μ
a. ∅	*!	
b. [[C <sub>μ</sub> V <sub>μ</sub> ] <sub>σ</sub> ] <sub>PrWd</sub>		*

This conclusion however only adds to the conundrum, since the opposite ranking DEP-μ >> MPARSE seems to be required for the monovocalic forms. To deal with this problem, recall that Bella Coola allows underlyingly long vowels, therefore any ban that needs to be imposed has to be on long vowels which originate from short vowels in the input. To capture this we could use the notion of old and new markedness (McCarthy 2003a). Within this model, the Bella Coola facts currently investigated present a case of a “grandfather effect”, where violations of a markedness constraint are tolerated if they come from the input, but not if they come from the output. This would mean that branching structures, like the ones occurring in long segments, are allowed if such branching (\*<sub>O</sub>BRANCHING) occurred in the input, but they are banned if they are first introduced in the output (\*<sub>N</sub>BRANCHING)<sup>27</sup>.

(29) \*<sub>N</sub>BRANCHING >> MAX-μ >> \*<sub>O</sub>BRANCHING

(30) V inputs - monomoraic and bimoraic: MAX-μ >> \*<sub>O</sub>BRANCHING<sup>28</sup>

V <sub>μ</sub>	* <sub>N</sub> BRANCHING	MAX-μ	* <sub>O</sub> BRANCHING
a. V <sub>μ</sub>			
b. V <sub>μμ</sub>	*!		
V <sub>μμ</sub>	* <sub>N</sub> BRANCHING	MAX-μ	* <sub>O</sub> BRANCHING
c. V <sub>μ</sub>		*!	
d. V <sub>μμ</sub>			*

Bella Coola underlying mono-moraic vowels surface as short vowels (30a). Effectively, branching is banned (30b), unless vowels are underlyingly branching. In this case, they will surface as long in the output as in (30d). Now all we need is to rank \*<sub>N</sub>BRANCHING above MPARSE, so that null outputs are a better solution when it comes to structures requiring newly introduced branching configurations in the output. MAX-μ >>

<sup>27</sup> An informal definition of \*<sub>N</sub>BRANCHING could be that branching in segments is not allowed [i.e. no long vowels and no geminates].

<sup>28</sup> Actually, as shown here, \*<sub>N</sub>BRANCHING could be ranked anywhere, but I simplify in presenting it top-ranked from the beginning, as the ranking arguments \*<sub>N</sub>BRANCHING >> MPARSE and MPARSE >> \*<sub>O</sub>BRANCHING in (32) immediately below indirectly suggest.

\*<sub>O</sub>BRANCHING on the other hand ensures that underlying branching structures will survive unaffected in the output<sup>29</sup>. With this modification, the correct result for monosegmental roots obtains. Note that monomoraic outputs are not considered since these will have already failed due to a WDMIN violation.

(31) \*<sub>N</sub>BRANCHING >> MPARSE >> \*<sub>O</sub>BRANCHING

(32) /V<sub>μ</sub>/ and /V<sub>μμ</sub>/ roots: \*<sub>N</sub>BRANCHING >> MPARSE >> DEP-μ, \*<sub>O</sub>BRANCHING

	V <sub>μ</sub>	* <sub>N</sub> BRANCHING	MPARSE	DEP-μ	* <sub>O</sub> BRANCHING
a.	V <sub>μμ</sub>	*!		*	
b.	∅		*		
	V <sub>μμ</sub>				
c.	V <sub>μμ</sub>				*
d.	∅		*!		

For a monomoraic input, the null parse wins, since the alternative, namely the lengthened vowel incurs a violation of \*<sub>N</sub>BRANCHING. Most significantly, observe that (32a) establishes the necessity of \*<sub>N</sub>BRANCHING; had this constraint been absent, then the ranking MPARSE >> DEP-μ alone would favor the lengthened candidate (32a) instead. In the case of an underlying bimoraic input though, the null output fails since the bimoraic winning candidate only violates the low-ranking \*<sub>O</sub>BRANCHING<sup>30</sup>.

<sup>29</sup> This analysis is not without problems either. Apart from some general difficulties that comparative markedness carries with it (Yip 2003), there are also some additional language-specific issues that need to be considered. The problem is how \*BRANCHING is to be understood. Although its interpretation as an output constraint is straightforward, the input representation it requires appears troublesome. The reason is that it entails underlying branching. Bimoraic vowels will have two moras in the input, but their links may be a product of output considerations. Instead we are here forced to say that the links are present in the input. This leads us back to enriching inputs, a task we have been trying to avoid (although recent work on geminates argues that underlying branching is perhaps necessary, cf. Muller 2001, Curtis 2003). Notice however that comparative markedness may not be the only way out of this problem. An alternative is the use of constraints such as MAX/DEP-LINK-μ after Morén (1999/2001) that ban the deletion or insertion of moras linked to specific segments (vowels or consonants). This approach however seems to bear on the same assumption, i.e. that there are underlying links. The reason I choose Comparative Markedness here is for expository reasons only. It requires the addition of fewer constraints and can be illustrated in the absence of links (this does not mean that these are not there though).

Finally, note that there is an alternative that does not make reference to link structure. This would be the constraint \*LENGTHEN which would penalize segments that get lengthened in the output and not those that are already long. By ranking \*LENGTHEN >> DEP-μ, the correct results would obtain. MAX-μ would not interfere, thus the ranking we required all along, i.e. \*LENGTHEN >> MPARSE >> DEP-μ would be possible. The problem with this view however is that \*LENGTHEN describes a process rather than expressing a representation that evaluates an output configuration.

<sup>30</sup> For consonants, the comparison would be one between /C/ and /C<sub>μ</sub>/, since long consonants are assumed to have branching but only one mora. For both inputs however and irrespective of old and new branching considerations, the competing outputs would be the monomoraic branching C and the null parse. At any rate, the former would violate WDMIN and thus be excluded. Vowel epenthesis could apply too, but this can easily be prevented by high-ranking DEP-SEG.

So up to now we have seen that a series of constraints dominates MPARSE, i.e. RTMAX >> MPARSE (23), WDMIN >> MPARSE (27), DEP-SEG >> MPARSE (27), \*<sub>N</sub>BRANCHING >> MPARSE (32). WDMIN also dominates \*MORAIC ONSET (19). Apart from MPARSE's domination over DEP-μ (28), it is also the case that MPARSE dominates \*<sub>O</sub>BRANCHING (32). \*MORAIC ONSET also dominates \*<sub>O</sub>BRANCHING by implication since the following holds: \*MORAIC ONSET >> MAX-μ (20) and MAX-μ >> \*<sub>O</sub>BRANCHING (30). What this suggests then is that the relative ranking of MPARSE and \*MORAIC ONSET may be somehow related, but we cannot yet determine exactly how. The introduction of SLH below supplies the missing link.

Before moving on, recall that in monosegmental words there is no way that WDMIN and \*<sub>N</sub>BRANCHING can be simultaneously satisfied, thus the optimal output is found in the guise of the null parse. However, in CC or CV words, \*<sub>N</sub>BRANCHING violations can be avoided, and WDMIN can be remedied by either allowing unsyllabified moraic consonants or by having a moraic onset, rendering the null parse an impossible candidate. Such forms are discussed presently.

We already know that in a CC sequence, where both consonants are obstruents (TT), no nucleus can be formed. This can be accounted for by the ranking:

- (33) \*<sub>NUC/OBSTR</sub> >> <sub>NUC</sub> >> \*<sub>NUC/SON</sub> >> \*<sub>NUC/VOWEL</sub>

This expresses that only sonorants and vowels can serve as nuclei, but not obstruents. Given that DEP-SEG and \*<sub>NUC/OBSTR</sub> are very highly ranked, a /CC/ form is destined to either arise as the null output or unsyllabified as illustrated below (*for full-ranking and justification of \*<sub>NUC/OBSTR</sub> >> SLH see Appendix A*). Note that this is the first time that SLH (Strict Layer Hypothesis) must be included in the ranking.

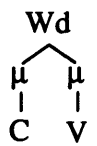
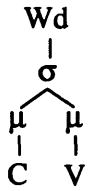
- (34) SLH: For every word, there is at least one mora dominated by a syllable

- (35) /C<sub>(μ)</sub>C<sub>(μ)</sub>/ --- [C<sub>μ</sub>C<sub>μ</sub>]<sub>PrWd</sub>: MPARSE >> DEP-μ, SLH

	* <sub>NUC/OBSTR</sub>	DEP-SEG	MPARSE	DEP-μ	SLH
a.				(*)(*)	*
b. ∅			*!		

SLH is violated by (35a) since both consonants remain unsyllabified and no syllable is formed. (35b) loses though as it presents the more serious violation of MPARSE. The question now is whether SLH can have any unwelcome repercussions in the consideration of our very first example, i.e. /CV/, which is clearly both a syllable as the reduplication facts show (§4.2.1) and has a moraic onset since it satisfies WDMIN. The answer is negative provided SLH >> \*MORAIC ONSET.

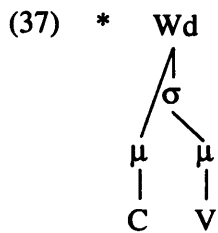
(36) /C<sub>(μ)</sub>V<sub>μ</sub>/ --- [[C<sub>μ</sub>V<sub>μ</sub>]<sub>σ</sub>]<sub>PrWd</sub>: DEP-μ, SLH >> \*MORAIC ONSET

	DEP-μ	SLH	*MORAIC ONSET
a. 	(*)	*!	
b. 	(*)		*

Candidates (36a) and (36b) are very similar, the only difference being that (36b) is syllabified. Evidence for syllabification has been supplied earlier in our discussion of reduplication in §4.2.1. SLH is decisive in favouring (36b) over (36a) by virtue of SLH >> \*MORAIC ONSET. SLH is not a high priority of the language, but whenever it can be satisfied, it will do so as it does here<sup>31</sup>. On the assumption of moraic inputs, nothing really changes apart from a few extra DEP-μ violations, which at any rate do not affect the outcome.

An additional reasonable candidate is the one below in (37) where the only thing different from the winning candidate is that the mora is not linked to an onset, but to an unsyllabified consonant. Apparently ONSET is in operation here, requiring that whenever syllabification can apply, the prevocalic consonant must syllabify as an onset instead of staying unsyllabified.

<sup>31</sup> The constraint ONSET will be used next. One would wonder whether this could be used here instead of SLH. I believe the answer is no. ONSET will impose certain requirements on prenucleic consonants of an already syllabified nucleus [as in (37)]. In (36a), there is no syllabification at all, so ONSET would be vacuously satisfied.



Thus, a further ranking is now established:

(38) ONSET >> \*MORAIC ONSET

	ONSET	*MORAIC ONSET
a. $[[C_{\mu}V_{\mu}]_{\sigma}]_{Wd}$ (36b)		*
b. $[C_{\mu}[V_{\mu}]_{\sigma}]_{Wd}$ (37)	*!	

The only remaining word-minimality case that we have not yet discussed is that of /VC/. Two facts are relevant here. The first one is that Bella Coola prefers syllabified codas to unsyllabified consonants. This suggests that the constraint which militates against unsyllabified segments, namely PARSE-SEG must dominate NO CODA as shown below.

(39) PARSE-SEG >> NO CODA

Moreover, codas are moraic in the language, hence MORAIC CODA >> DEP-μ. Taking this into consideration, alongside the MPARSE >> SLH established before in (35), the following ranking obtains (*relevant tableaux in Appendix A*):

(40) MORAIC CODA, MPARSE >> DEP-μ, SLH

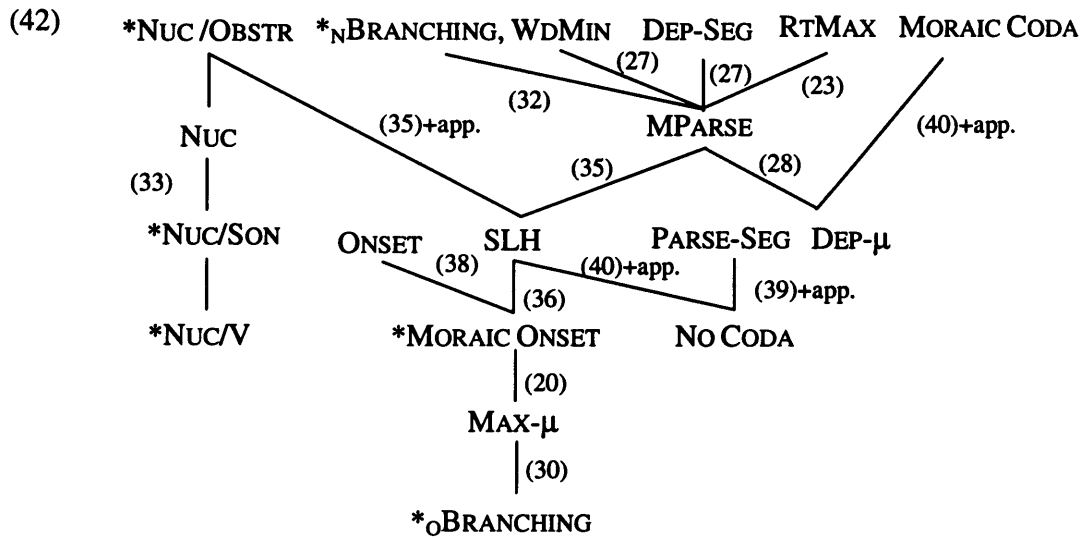
To sum up, Bella Coola monomoraic words are banned due to Word Minimality restrictions. Bimoraic or larger words are well-formed. In the case of bimoraic words, we can find those that end up unsyllabified, i.e. CC, with an obstruent cluster or those which are syllabified by virtue of a sonorant segment that can act as a nucleus, e.g. CV or VC. The latter form presents a straightforward example of a heavy syllable where the coda contributes to weight. Things are more surprising in the case of CV, which, as we know, has to be both syllabified and bimoraic. Since one mora comes from the vowel, the other is bound to come from the onset. This is the only example where we can see



moraic onsets in Bella Coola in action. Thus, the major constraints pertaining to Word Minimality and their respective ranking is given below.

(41) WDMIN >> MPARSE >> DEP- $\mu$ , SLH >> \*MORAIC ONSET

So the picture we have shaped so far looks like [App=reference to Appendix A]:



#### 4.4.2 Against moraic onsets

Before concluding the analysis of WdMin, I will briefly examine an alternative suggested by Elliott Moreton (p.c.) who claims that there is a way out without making reference to onset moraicity. Suppose that MAX-SEG dominated WDMIN and that we only admitted the standard notion of bimoraicity. Then all of CVC, VC, and VV words would be well-formed by virtue of bimoraicity, while CV ones would be accepted due to the overriding importance of MAX-SEG. C and V words however, despite satisfying MAX-SEG would fail minimality, thus they would be unable to surface. The problem with this approach is that there is no reason why /CV/ would not lengthen to [CVV] or [CVC] so that it would simultaneously satisfy bimoraicity.

One possibility is that DEP-SEG is also top-ranked so that /CV/ cannot become [CVC]. This would not stop it from becoming [CVV] however, so again no good explanation obtains. An apparent remedy would be the use of high-ranked DEP- $\mu$ , so that lengthening is banned. But DEP- $\mu$  cannot be highly-ranked, because we want to allow for the addition of moras on syllables with codas or on unsyllabified consonants. It seems that employment of \*<sub>N</sub>BRANCHING bypasses this problem in the way shown

earlier too (cf. (32)). By having it highly-ranked, /CV/ → [CVV] would be banned without affecting /CVC/ → [CV<sub>μ</sub>C<sub>μ</sub>]. Is then the ranking in (43) a real problem for my analysis? Consider the following tableaux that illustrate the problem:

(43) \*<sub>N</sub>BRANCHING, MAX-SEG, DEP-SEG >> DEP-μ, WDMIN >> MPARSE

	/V <sub>μ</sub> /	* <sub>N</sub> BRANCHING	MAX-SEG, DEP-SEG	DEP-μ	WDMIN	MPARSE
a.	V <sub>μμ</sub>	*!				
b.	V <sub>μ</sub>				*!	
c.	Ø					*
	/V <sub>μμ</sub> /					
d.	V <sub>μμ</sub>					
e.	Ø					*!
	/C/					
f.	CV <sub>μ</sub>		*!			
g.	C				*!	
h.	Ø					*
	/CV/					
j.	CV <sub>μμ</sub>	*!		*		
i.	CV <sub>μ</sub>				*!	
k.	Ø					*

This grammar yields the right results for /V/ and /C/ roots where the null parse is generated (43c, h), while it chooses a long-voweled output for a /V<sub>μμ</sub>/ input. The problem arises in a /CV/ input where the optimal candidate is wrongly the null parse (43k). The problem could be resolved if the ranking between WDMIN and MPARSE was reversed. MPARSE >> WDMIN would now select (43i) as the winner, but it would also entail that [V] and [C] outputs [(43b) and (43g) respectively) should survive too. This is clearly wrong as the empirical facts show. Thus, an alternative which makes no reference to onset moraicity is not viable. With this in mind, we can consider the rest of the facts.

#### 4.4.3 The remaining roots

As already mentioned, Bella Coola allows numerous unsyllabified consonants, which are not stray erased. This can be explained using the notion of moraic licensing after Bagemihl (1991, 1998) and Lin (1997) for Piro. In these works, it has been proposed that unsyllabified consonants do not delete because they can be licensed by moras. This is an assumption shared here as well.

- (44) MORAIC LICENSING (MLIC): An unsyllabified consonant is realized, i.e. licensed, when dominated by a mora

We know that the maximal syllable in Bella Coola is (C)CVVC<sup>32</sup>, which means that complex onsets are allowed (provided they are of the obstruent-sonorant type), long vowels are permitted, but complex codas are banned. Evidently \*COMPLEX CODA is undominated in the language. Given the maximal syllable, any remaining surface obstruents must be unsyllabified, but also be moraic so that they satisfy MLIC. At the same time, another constraint, MAX-SEG requires that as many segments as possible surface. The obvious solution to reconcile MAX-SEG which requires the output realization of all segments and MLIC which deletes non-moraic unsyllabified consonants is to attach a mora to all consonants. In this way, both constraints are satisfied.

Under this view, MLIC is seen as a trigger for mora insertion and therefore has to dominate the constraint which punishes such an insertion, i.e. DEP- $\mu$ . Notice however, that mora insertion occurs, but is not uncontrolled. It is restricted by RTMAX. The next sub-sections highlight how MLIC operates in well-formed (§4.4.3.1) and maximal (§4.4.3.2) roots in combination with other constraints.

#### 4.4.3.1 Well-formed roots

I will briefly examine how the analysis built up to now accounts for roots which are neither minimal nor maximal, but somewhere in between. First, recall that we have seen moraic onsets only emerging in CV words. This was due to a combination of constraints and of a particular ranking, but in larger roots, WDMIN can be satisfied by other means, thus violations of \*MORAIC ONSET prove gratuitous, as we have seen in the /CVC/ case in (20).

Now we will examine how unsyllabified consonants are dealt with. So far and with the exception of CC clusters, we have not seen exactly how these are treated. To express their retention on the surface, MLIC must be introduced in the tableau. The minimally larger /CVCC/, e.g. *sutk* “winter” (B98: 75) serves well for this purpose.

---

<sup>32</sup> We have already seen an example of the CVVC type, e.g. *si.taax.su*. The claim is that the language posits a bimoraic maximum for each syllable, therefore in the syllable *taax*, the two moras come from the vowel, while the coda bears none.

(45) /CV<sub>μ</sub>CC/ --- [[CV<sub>μ</sub>C<sub>μ</sub>]<sub>σ</sub>C<sub>μ</sub>]<sub>PrWd</sub>: MLIC >> DEP-μ

CV <sub>μ</sub> CC	MLIC	DEP-μ
<p>a.</p>		**
<p>b.</p>	*!	*

Bearing in mind that the RtMax data establish that unsyllabified consonants carry a mora, we only need examine the two candidates above to establish that MLIC >> DEP-μ. Numerous other candidates can be considered for this input. The interested reader may find these as well as discussion of the moraic input in *Appendix A*. Candidate (45b) is identical to (45a), the only difference being that it fails to assign a mora to the unsyllabified consonant at the end of the root. This proves fatal, because of the MLIC violation.

The final part of the analysis explores cases with larger roots and shows the effects of MAX-SEG and RTMAX. The latter expresses the restriction examined in section §4.2.2.1 regarding the four mora maximum per root. A few additional but more peripheral constraints will also be discussed to complete the analysis.

#### 4.4.3.2 Larger roots and root maximality

The RTMAX constraint permits us to discard numerous segmental sequences which are not predicted to occur, and indeed they do not, such as those below:

(46) Roots exceeding the 4μ-maximum<sup>33</sup>

*Disyllabic roots*

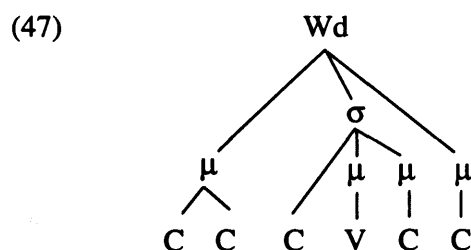
a. CVCCCCV	*	5μ	[μμ]μμ[μ]
b. CVCCCCCV	*	6μ	[μμ]μμμ[μ]
c. CVCCCVC	*	5μ	[μμ]μ[μμ]

<sup>33</sup> Recall that boldface is used for the unsyllabified consonants.

d. CVCCCCVC	*	6μ	[μμ]μμ[μμ]
<i>Trisyllabic roots</i>			
e. CVCCVCVC	*	5μ	[μμ][μ][μμ]
<i>Monosyllabic roots</i>			
f. CCCVCC	*	5μ	μμ[μμ]μ

Based on the observations made above, e.g. that the maximal syllable is TRVVC, where T=obstruent, R=sonorant, as well as that complex codas are banned, while extra consonants may be unsyllabified, these are the facts that we indeed expect. In all these roots, unsyllabified consonants are utilized, but at the same time, the maximum morality is exceeded, hence these forms do not after all make it to the surface.

Observe however that RTMAX merely says that roots should have no more than four moras. As a result, the analysis produced so far can only exclude the candidate \*C<sub>μ</sub>C<sub>μ</sub>CV<sub>μ</sub>C<sub>μ</sub>C<sub>μ</sub> for an input like /CCCVCC/ as it includes five moras. We know that eventually for such cases, the null parse wins, but rather than producing an output that does not realize any segments, other reasonable candidates can be considered all of which satisfy RTMAX and delete one mora to obtain the four-mora maximum. The main contenders are the following: a) (47) with mora sharing, b) (49) with segment deletion and c) (51) with unsyllabified non-moraic consonants.

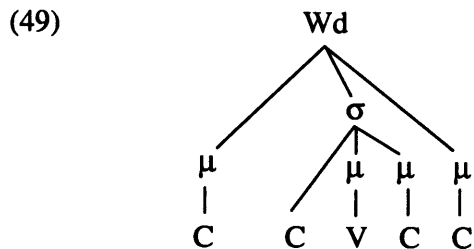


Here the onset is non-moraic, while the nucleus V and the coda C receive a mora. There are three unsyllabified consonants, but because only two more moras are available, two of the segments share a mora. Such a representation violates \*MORA SHARING<sup>34</sup>.

<sup>34</sup> Bagemihl does not consider representations like (47), and makes no reference to the ban against mora sharing. Nonetheless, in (47), the first two consonants share a mora and are in this sense licensed. Thus, in his system, (47) is left unaccountable for, unless he implicitly assumes a formulation of moraic licensing that requires each segment to be licensed by a unique mora, so that mora sharing is not allowed. At any rate, this is not discussed and remains a weak point in his approach.

Furthermore, note that in a different context, Bagemihl seems to allow mora sharing although perhaps unintentionally. In his example (1) [B98: 73], the syllable *ʔan* is naturally parsed in a CVC syllable. The coda has a mora and the vowel has one which is also shared with the onset consonant. This constitutes mora sharing and provides an extra argument that a representation like the one in (47) is left unaccounted for in his model.

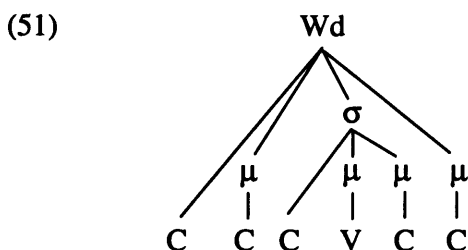
- (48) \*MORA SHARING: Segments do not share moras (Sprouse 1996: 398, 406; Broselow, Chen and Huffman 1997: 65)<sup>35</sup>



This time, instead of mora sharing, a consonant has deleted, violating MAX-SEG<sup>36</sup>.

- (50) MAX-SEG: Do not delete segments

Finally, it is possible to leave one of the unsyllabified consonants without a mora. This causes a moraic licensing violation since every unsyllabified segment must be dominated by a mora.



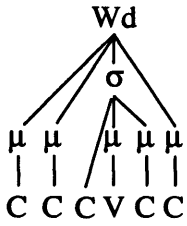
Three consonants are unsyllabified, but unlike the other two, the first one is also non-moraic leading to an ill-formed structure. Again no difficulty emerges in accounting for this example, because as we know already, MLIC is top-ranked. All of the above can be summarized in the following tableau:

<sup>35</sup> The word *sitaaxsu* in (11f) could be seen as an instance of mora sharing, since the moras contributed by the vowels are already four. Perhaps then the consonant *χ* shares a mora with the preceding vowel. This would obviously violate \*MORA SHARING. In the absence of compelling evidence, we could however claim that the widely-attested \*3μ constraint is in operation and forces the coda consonant to surface as non-moraic. Given that the consonant is syllabified, this does not present any problems in terms of moraic licensing.

<sup>36</sup> This possibility cannot really be ruled out. Perhaps the input for the word *sqʷaʔkʷ* “ashes” (B98: 76) may be the hypothetical /tsqʷaʔkʷ/. Then the initial consonant could delete and we would obtain the output above. However, by Lexicon Optimization (Prince and Smolensky 1993/2004), the input /sqʷaʔkʷ/ will be preferred. Moreover, since RTMAX is a root constraint, all the forms it evaluates are underived. As a result, no alternations - occurring in processes such as deletion in derived environments - emerge that could offer us empirical evidence towards one direction or another.

(52) /C<sub>(μ)</sub>C<sub>(μ)</sub>C<sub>(μ)</sub>V<sub>(μ)</sub>C<sub>(μ)</sub>C<sub>(μ)</sub>/ --- Ø<sup>37</sup>

RTMAX, MLIC, \*MORA SHARE >> MAX-SEG >> MPARSE >> DEP-μ

	RT MAX	MLIC	*μ SHARE	MAX- SEG	M PARSE	DEP -μ
a. 	*!					(*)(*) (*)(*)
(51) b. [CC <sub>μ</sub> [CV <sub>μ</sub> C <sub>μ</sub> ] <sub>σ</sub> C <sub>μ</sub> ] <sub>PrWd</sub>		*!				(*)(*) (*)
(47) c. [C <sub>μ</sub> C <sub>μ</sub> [CV <sub>μ</sub> C <sub>μ</sub> ] <sub>σ</sub> C <sub>μ</sub> ] <sub>PrWd</sub>			*!			(*)(*) (*)
(49) d. [C <sub>μ</sub> [CV <sub>μ</sub> C <sub>μ</sub> ] <sub>σ</sub> C <sub>μ</sub> ] <sub>PrWd</sub>				*!		(*)(*) (*)
e. Ø					*	

Candidate (52a) has five moras and thus exceeds the four mora maximum; (52b) includes an unsyllabified non-moraic consonant that violates MLIC. The next candidate presents mora sharing, while (52d) deletes a segment which causes a MAX-SEG violation. (52e) therefore wins since it incurs the least severe MPARSE violation. The same results would occur even if the input were moraic, since all the violated constraints are markedness constraints. DEP-μ is too low ranked to affect the outcome in any way as is shown by the violations within brackets in the last column.

In sum, whenever the root exceeds the four mora maximum imposed in the language, the form cannot survive, since it violates RTMAX. Other possible forms are excluded too, because these also violate other high-ranking markedness constraints. Consequently, the null parse proves to be the optimal output.

#### 4.4.4 Summary and Final Ranking

In this chapter, I have offered an analysis of Bella Coola roots. Although the focus has been on word minimality, root maximality has also been considered to provide a fuller

<sup>37</sup> It is possible to have MAX-SEG among these top-ranked constraints, i.e. RTMAX, ..., MAX-SEG, but this also permits the interpretation MAX-SEG >> RTMAX, which is excluded, since it is never the case that roots exceed the mora maximum in order that more segments are realized. To avoid this problem, I present MAX-SEG dominated, but still highly-ranked, since given the chance - when the other highly-ranked constraints are satisfied - the highest possible number of segments is realized.

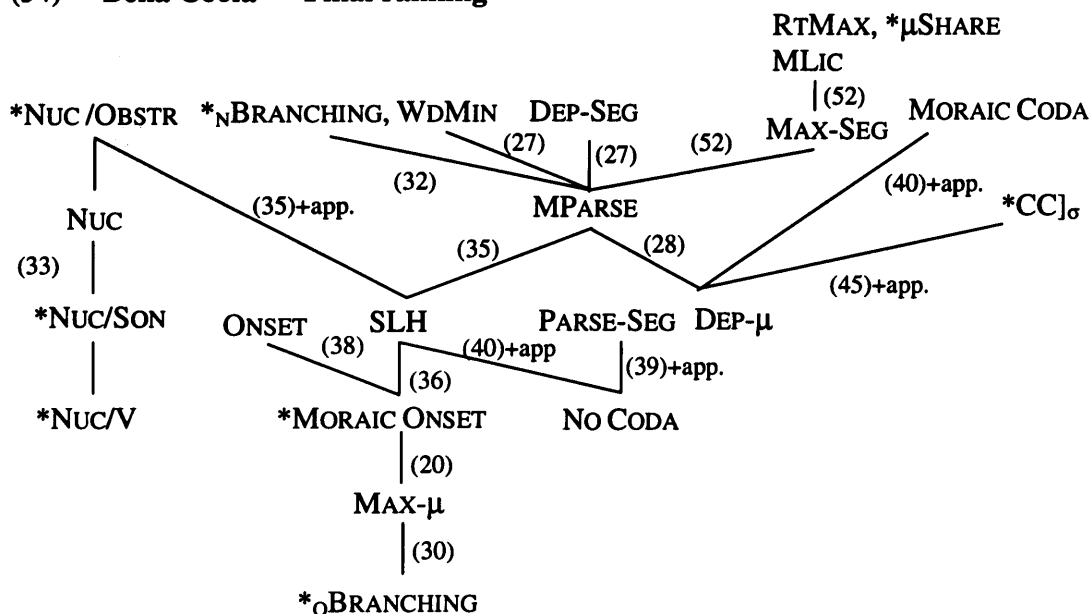
account of the language's roots and establish that moraic onsets are only present in CV words. The core phenomenon of Word Minimality is just described by:

(53) **WDMIN >> MPARSE >> DEP- $\mu$ , SLH >> \*MORAIC ONSET**

WDMIN is a very highly ranked constraint requiring that all words satisfy it. In CV words, this leads to the emergence of moraic onsets. At the same time, no segments can be added in the output, whereas the addition of moras is limited. This means that for words made of a single vowel or consonant, only the null parse can be generated. In words of the CC type, the SLH is violated, since these consonants cannot be syllabified. However, these carry moras and fulfil the word minimum despite the lack of syllabification.

Outside Word Minimality and more generally in the language, the preferred syllabification is quite typical in that CVC syllables are constructed where the nucleus is either a vowel or a sonorant. Complex onsets of rising sonority are also allowed, but complex codas are banned. Frequently, obstruent consonants cannot syllabify, but do not delete either, because MAX-SEG is high-ranked. The only way that these consonants can surface is by being moraically licensed (high-ranking MLIC). Indeed, this is what happens, provided that the total number of moras within a root does not exceed four moras (due to high-ranking RTMAX). If the maximum is exceeded, the null parse wins. The resulting ranking for Bella Coola is:

(54) **Bella Coola --- Final ranking**





The current analysis offers an accurate representation of the Bella Coola facts, but is also advantageous over Bagemihl's (1998) proposal in several respects. Making use of the OT framework, I have provided an account that dispenses with two undesirable assumptions of Bagemihl's, namely crucial reference to underlying moraicity and satisfaction of Word Minimality in the input. The present account is instead fully output-oriented. Moreover, due to the inherent violability of the OT constraints, it is possible to express that moraic onsets will only arise in a restricted environment, only as a means to satisfy WDMIN. When this is not at stake, \*MORAIC ONSET which is still active, albeit low-ranked, ensures that no superfluous weightful onsets will arise.

Before concluding, I would like to stress that the rankings presented in the preceding analysis are entirely consistent with the empirical facts, but some technical details could be somewhat different from what has been sketched out here. In particular, I have argued that the null parse is the winning candidate for /V/ and /C/ roots as well as for roots that exceed the mora maximum imposed in the language. It is however logically possible that /V/ and /C/ roots may actually lengthen to [VV] and [CC] or epenthesize a segment so that /C/ becomes [CV]. Similarly, extra-large roots could shorten by deleting segments and moras to fit the RtMax pattern.

By looking at the empirical facts we simply cannot tell what exactly happens, but even if it were the case that these alternative strategies were optionally or wholly adopted, then we would just require some modification of the rankings suggested. For example, in (27), minimally altering the ranking from WDMIN, DEP-SEG >> MPARSE to WDMIN >> DEP-SEG, MPARSE would generate both the null parse and [C<sub>μ</sub>V<sub>μ</sub>] as possible outputs for a /C/ word. Similar alterations could apply in other cases too.

In the absence of empirical confirmation towards one direction or the other, I have chosen to provide a grammar with stricter rankings than the ones required in the alternative cases. Note also that examination of some general facts of the language suggests that re-locating certain constraints might actually not be so desirable. For instance, in the above ranking, DEP-SEG has been demoted so that it can be violated in the language. This would imply that epenthesis is more generally allowed in the language.

However, Bagemihl only mentions schwa insertion as a transition between an obstruent and a following sonorant consonant regardless of the syllable structure, thus its insertion does not relate to the presence of unsyllabified segments. Bagemihl claims that such a schwa is actually excrescent rather than epenthetic, since it bears all the characteristics of excrescent vowels. Its quality is variable and subject to the

surrounding environment, while it is not referenced by any morphological or phonological processes (B91: 600). If it participated in morpho-phonology, then for instance, the base *t/k<sup>w</sup>* after schwa insertion would become *talk<sup>w</sup>* and should reduplicate as *\*tətalk<sup>w</sup>* which is incorrect. In fact, Bagemihl claims that “no other process of epenthesis is reported for Bella Coola” (B91, fn. 12), a fact that would run counter to the modified ranking suggested above.

Further exploration into the language would thus be required to establish a grammar that would take into account all the other possibilities, but remain consistent with the general Bella Coola facts. For the time being, the grammar proposed in (54) achieves the desirable effects and does not bear any negative repercussions of the sort described.

In my view, however, the most important point is that irrespective of the specific details of the grammar, nothing changes with respect to the - up to now - controversial [CV] facts. No matter what the source of [CV] outputs is, e.g. /V/, /C/ or /CV/, reference to moraic onsets must be made.

## 4.5 Residual issues

### 4.5.1 Morphological Root vs. Morphological Word

In the preceding discussion, it has been made clear that the restrictions relevant to Maximality refer to the root. No such restrictions apply at the morphological word level, as the following examples reveal (B98: 74):

- (55) *c'klaktk<sup>w</sup>p* “ten fathoms” (*c'klakt* “ten”, *-k<sup>w</sup>p* “fathoms”)  
*łx<sup>w</sup>tłcx<sup>w</sup>* “you spat on me” (*łx<sup>w</sup>t* “to spit”, *-ł* PAST TENSE, *-c* 1SG, *-x<sup>w</sup>* 2SG)

This may be problematic for an output-oriented theory as it may first presuppose syllabification and moraification at the root level, followed by syllabification and moraification at the word level, suggesting a serial derivation. This however may not be necessary as is argued immediately below. Much of what follows though is of a more speculative kind due to the lack of sufficient data that would test the current proposal or other alternatives.

First, we have to notice that even if the syllabic and moraic behaviour across morphemes differs from the intra-morphemic one at the root level, this may not

necessarily condemn the current approach or indeed any OT approach. It is not at all random that such potential difference coincides with morpheme boundaries. Within OT there have been descendants of Lexical Phonology (e.g. Kiparsky 2000), that propose different constraint rankings depending on the morphological level evaluated each time.

Another possibility, which I am personally more inclined to adopt is the necessity of constraints that refer to certain morphological levels. Downing (1999) for instance shows the need to make reference to a particular morphological constituent to account for processes such as reduplication or infixation in Bantu languages. More concretely, the reduplicant in Kikuyu and Kinande consists of a disyllable that copies the first foot of the base. The interesting thing is that the final vowel of the reduplicant is always the vowel /-a/ irrespective of the root's vowel. Although the choice of /a/ is phonologically arbitrary, morphologically it is unmarked, since it is the ending of the disyllabic canonical Bantu verb stem. Downing concludes that not only metrical constituents may serve as prosodic templates, but also morphoprosodic constituents such as the canonical, i.e. in this case disyllabic, verb stem may do so.

Bagemihl (1998: 89) also cites work by Orgun and Inkelas (1992) where minimality constraints in Turkish may be restricted to certain sub-lexical levels instead of applying to the entire lexicon. Even more convincing evidence comes from Musqueam, which like Bella Coola is a Salish language (Shaw 2002). Shaw examines obstruent clusters in this language and observes that certain restrictions may apply at the morphological word domain and not at the morphological root one or conversely. To be more specific, Place and Manner restrictions apply at the MWd but not at the MRt domain. On the other hand, schwa-epenthesis occurs to break up obstruent clusters within the root, but not within the word. Finally, the maximal initial obstruent sequence at the word domain is [OOO+], whereas for the root the restriction is [OO. Additionally, Shaw (2004) shows that stress consistently falls on the first syllable of the morphological root. In the light of the above, she concludes that the morphological root constitutes a distinct prosodic domain that has to be recognized at the output level.

The current proposal is along these lines, since it admits a constraint that makes specific reference to the maximality of the root. Bella Coola is now added to the pool of languages that provide support that the morphological root may serve as a distinct prosodic domain.

#### 4.5.2 MPARSE

The current analysis makes use of the constraint MPARSE (Prince and Smolensky 1993/2004, McCarthy and Wolf 2005) to deal with candidates that do not manage to receive any phonetic representation. However there are other conceivable accounts for this problem which will be reviewed here.

Fanselow and Féry (2002) refer to a number of examples drawn from phonology, morphology and syntax and argue that parochial language-specific constraints are unavoidable. They claim that such constraints could obviously be incorporated in EVAL, so that in languages that they have an effect on, they would be highly-ranked and as a result play a decisive role in the output. Conversely, in the languages where they have no effect at all, they would be so low-ranked that practically they would be inactive. The problem with such a view is that it would run the risk of making OT non-falsifiable, since for each construction, there would be some (parochial) constraint that could account for it.

Instead, what they propose is that in addition to EVAL, there is also a separate control component that includes language-dependent constraints or morpheme-specific restrictions. Under the proposal that Control follows EVAL, candidates may be optimal through EVAL but parochial constraints or lexical restrictions of Control (e.g. "goed" cannot survive since "went" already exists) may block them from appearing. Turkish and Chaha are the relevant examples they offer that somehow resemble Bella Coola. Turkish has a WdMin requirement which cannot always be satisfied. If the available repair strategies cannot apply, i.e. epenthesis is blocked by hiatus, then the form fails to surface. In Chaha, frequentative verbs must consist of four consonants, so in triliteral roots the second consonant is copied to fill the extra C slot, but there are no frequentative forms for diliterals as double-copying is banned.

Under this approach, in Bella Coola where roots are maximally quadrimoraic, forms that exceed this number are impossible since on the one hand moraic licensing requires all unsyllabified consonants to be moraic, so that no consonants can remain stray and on the other hand no consonants can be deleted or be allowed to share moras. Since all these cannot be simultaneously satisfied, no output can be generated.

Nonetheless, there are two important reasons not to adopt this idea for at least the Bella Coola data. The more general reason is that, as the authors admit, there is not yet an explicit mechanism of how the control component is organised. As a result, we cannot articulate a full proposal for the Bella Coola large-root cases, contrary to the

MPARSE analysis which can be directly incorporated to the general schema proposed for the language. In addition, Bella Coola does not seem exactly analogous to the cases mentioned in Fanselow and Féry, in the sense that - as we have seen in the previous section - there are other languages, e.g. Musqueam, which support the idea of allowing constraints that make reference to the maximality of roots. Therefore, the parochiality of *\*goed*, because of the lexical blocking of *went*, seems very different from the situation appearing in Bella Coola. By proclaiming such maximality restrictions 'language-specific', we may also lose the insight that there are certain languages that actually share such restrictions, and presumably these are not as language-specific as put forward in Fanselow and Féry (2002).

A similar but more promising approach is presented in Orgun and Sprouse (1999). They too propose a separate Control component, but this has different properties. First, they present convincing evidence that in some languages the MPARSE approach is not tractable. For instance, considering again the Turkish example, the word minimality condition cannot be satisfied by subminimal words, as these are not repaired by epenthesis, therefore no output obtains. This would suggest the ranking DEP >> MPARSE. At the same time, epenthesis is available for other processes, such as complex clusters or vowel hiatus avoidance. In this case, an output is indeed possible, thus MPARSE >> DEP. Similar examples are presented illustrating such a conflict.

The solution Orgun and Sprouse (1999) propose is that in fact EVAL as we know it in classic OT not only includes rankable violable constraints, but also inviolable ones. These are the ones that are never violated in a language, e.g. ONSET in a language where all syllables are or become through epenthesis, onsetful. Crucially though, a further distinction needs to be made; apart from the inviolable constraints that trigger a particular repair so that they remain unviolated (inviolable-in-practice), e.g. ONSET in the example above, there are also inviolable constraints which do not allow any repair at all (inviolable-in-principle). These only evaluate whether a candidate is grammatical or not. The way this works is the following: candidates are evaluated in the normal fashion. Then the proclaimed winner goes through the Control component which includes a constraint that evaluates the form's grammaticality. If this constraint is satisfied, then the form is rendered grammatical, otherwise ungrammatical.

Under this conception, in the case of Turkish above, DEP will be low-ranked in EVAL and dominated by constraints - which are inviolable-in-practice - against complex coda clusters or vowel hiatus, i.e. *\*V.V.*, *\*CODA COND* >> DEP, so that epenthesis may apply to avoid such violations. The winners of this evaluation are *then*

submitted to Control that includes the Word Minimality constraint and this will decide if the form may after all survive or not. Under this EVAL ranking, the optimal forms for words like /do-m/ “my C (musical note)” and /it-m/ “my dog” are [dom] and [itim]. But when these are submitted to Control, only the latter will pass through it, since the former violates Word Minimality.

Indeed, Orgun’s and Sprouse’s (1999) approach has a number of advantages, both empirical, e.g. accounting for problematic examples under the MPARSE model and conceptual, e.g. effectively and directly distinguishing inviolable constraints (for detailed argumentation the reader is referred to the original). However, a number of considerations are simultaneously raised. For example, the authors focus on the role of Control in systems where an output is judged ungrammatical. The question is then: what happens in a system where there is always an available winner? Would Control still need to be induced? And if so, how exactly would we know which constraint would belong to that component? Assuming that some constraint has to constitute this component, then presumably it should be one that is never violated in the language, but which? The obvious answer is that it would be one of the undominated constraints in EVAL. But as we have seen, undominated constraints in EVAL are those which trigger a repair to the output. On the other hand, Control contains constraints that do not allow repairs. In other words, for these systems, EVAL and Control should contain a constraint which is simultaneously inviolable-in-practice and inviolable-in-principle, but this seems impossible given Orgun’s and Sprouse’s model.

Even if a remedy can be thought for this problem and assume that there is such a constraint that fulfils these requirements, another issue is brought up, namely duplication. For instances where a unique winner is produced, i.e. the majority of cases, the same constraint would have to be considered in both the EVAL and Control components and this is obviously superfluous, loses insight and adds unnecessary complexity to the grammar. Finally, a practical consideration also arises. Which of the undominated EVAL constraints would be chosen as the Control constraint? So for instance, what if the language in question was like Turkish in having the two top-ranked constraints \*V.V and CODA COND (and apparently more when considering the whole grammar of the language), but different in that an optimal winner would be always produced for all cases? Would then \*V.V or CODA COND be the Control constraint?

I conclude that despite the advantages of the Control analysis of Orgun and Sprouse, a series of unresolved issues persist which make this approach difficult to

sustain. An MPARSE analysis works well and less controversially for the Bella Coola data at least, therefore it is the one I adopt.

## **4.6 Conclusion**

The present chapter has focused on a clearly weight-based phenomenon, that of Word Minimality and has shown that languages like Bella Coola calculate the onset for minimality purposes. Given that no re-analysis based on prominence is available for this phenomenon, we thus have a strong argument for the existence of moraic onsets. The introduction of weightful onsets in Bella Coola also eliminates the problems that the previous analysis of the same data by Bagemihl (1998) faced. Since the current account is fully output-oriented, it no longer treats Word Minimality as an input condition nor does it consider all segments as underlyingly moraic. Both assumptions were crucial in Bagemihl (1998), in order that CV words could satisfy the bimoraicity criterion of Word Minimality, but surfacing as monomoraic light syllables.

The present proposal instead argues that onsets are moraic, albeit in a very marginal environment, namely that of a /CV/ word. In this way, the overarching requirement of bimoraic Word Minimality can be satisfied. Elsewhere onsets are non-moraic, as the Root Maximality facts verify (cf. §4.2.2.1). Bella Coola thus presents a case of coerced onset weight, where onset moraicity is enforced by a higher imperative in the language, that of Word Minimality.

## Chapter 5

### CL from onset loss: the case of Samothraki Greek

*Μια φουά κι ένα κιό μες τα βάαγια τς Σαμαθάακς (Andriotis 1926-1928)*

*[Mja fu(r)á kjéna ki(r)ó mes ta v(r)áaxja ts Samaθ(r)áaks]*

*‘Once upon a time in the rocks of Samothraki...’*

#### 5.1 Introduction

In this chapter, I examine a paradoxical instance of compensatory lengthening (CL), that occurring in the Northern Greek island of Samothraki. It is paradoxical, because it apparently involves lengthening after onset loss, a situation which is held to be impossible among theorists, especially in Hayes’ (1989) seminal paper on CL. Instead I argue that such kind of lengthening is indeed possible. The fact that CL after onset deletion is so rare has other causes, such as the cross-linguistic observation that onsets delete much less often than codas as well as the fact that onset-deletion with CL (at least in simplex onsets) would lead to sequences such as VVV or V:V which are frequently disallowed. These two factors, also need to be considered in conjunction with the fact that even coda deletion does not always lead to CL. Sometimes, the consonant deletes without any remnant by means of lengthening. The standard explanation for that would be that such codas are non-moraic, hence their deletion may not lead to lengthening. This is not an adequate explanation, since there also exist cases where a vowel deletes, but no lengthening occurs either. One cannot attribute this to the lack of moraicity, because it is uncontroversial that vowels are always moraic. This issue is addressed in more detail in §5.7.

The general idea here is that CL is not treated as mora preservation, but rather as preservation of the position of a segment via mora addition. The major advantage of this approach is that it frees CL from intermediate and serial derivational levels. It can apply just as well to coda and onset deletion cases, while it makes the important prediction that some languages might even show CL from consonants which stay unsyllabified, as for instance in Piro (Lin 1997).



However, since the primary goal of this work is not compensatory lengthening, the proposed analysis should not be construed as a remedy for all cases of compensatory lengthening, but as a starting point for further investigation of CL as a whole. My attention is mainly drawn on CVC CL as Kavitskaya (2002) calls it, that is when a consonant deletes and causes lengthening. No account for CVCV CL is attempted at this stage. Notably, the current proposal remains phonological like Hayes' one (1989), instead of transferring much of the burden onto phonetics as in Kavitskaya (2002).

The chapter is structured as follows: I first briefly consider Hayes' (1989) CL account in a serial derivational framework (§5.2). This serves as the point of departure to show that a similar analysis cannot be maintained in an OT parallelistic framework for reasons to be discussed. I then present the current analysis and show in which way it is advantageous (§5.3). Having considered a few cases of standard CL, I move on to consider Samothraki Greek CL in large detail (§5.4). After the completion of that analysis, I examine possible alternatives and comment on the problems each faces (§5.5). Next, I explore the consequences of the proposed model and show how some cases which are unexplained under more traditional approaches, e.g. CL in a language that lacks syllable weight contrast and has unsyllabified consonants (cf. Piro), are straightforward here (§5.6). Finally, I offer some ideas on why onset-deletion CL is vanishingly rare (§5.7) and present some concluding remarks (§5.8).

## **5.2 Hayes' (1989) model and problems for OT**

Compensatory lengthening is a widespread phenomenon under which a segment (usually) deletes and as a result a neighbouring segment lengthens to compensate for its loss. Hayes has carefully investigated numerous instances of CL and concluded that a straightforward explanation can be given within a moraic phonology model. Working in a serial framework, he proposes that an input string first gets to be syllabified and simultaneously moraified. Only nuclei and codas may bear moras, since the assumption is that onsets never carry them. Then a rule applies that deletes a segment from a certain syllabic position. If this happens to be the coda - which can carry moras - then a neighbouring segment lengthens to cover for the former's loss. This is only possible because the deleted segment was in a position that can host a mora. Had it deleted from an onset position, CL would never occur, since onsets never host any moras.

Crucially, this process happens in stages, i.e. first taking an input like /kan/, then assigning moras and syllables [ $ka^{\mu}n^{\mu}$ ], then deleting an offending segment [ $ka^{\mu}\emptyset^{\mu}$ ] and

finally re-associating the mora left behind to a neighbouring segment [ $ka^{\mu\mu}$ ]. In this process, the use of the mechanism dubbed Weight-by-Position (WBYP) which ascribes moras on codas proves indispensable. Of course, WBYP must apply before segment deletion applies so that its effects by means of lengthening are visible.

While most cases of CL are successfully and straightforwardly accounted for in this model, it is notoriously difficult to translate this analysis to an output-oriented model such as OT. The reason is that since in classic OT only two levels are allowed, namely the input and the output, no level exists to allow intermediate syllabification and WBYP application, the two defining features of the Hayesian approach.

This makes CL a difficult issue in OT: if there are no intermediate stages, then how can lengthening occur after a segment deletes since there was never a mora that would cause such lengthening in the first place? One way to produce CL then is to have the mora that WBYP previously introduced be specified already in the input. This is exemplified below.

(1) CL in OT --- moraic input

	/CV <sup>μ</sup> C <sup>μ</sup> /	NoCODA	MAX-μ	MAX-SEG	DEP-μ
a.	CV <sup>μ</sup>		*!	*	
b.	CV <sup>μμ</sup>			*	

To simplify things, I am here using NoCODA so that the final C deletes (DEP-SEG is high-ranked banning vowel insertion as an alternative remedy). Both candidates under consideration fare equally well in terms of the remaining constraints. The only difference is that the winner, i.e. the lengthened candidate, satisfies MAX-μ perfectly by preserving all input moras in the output. We thus generate the right CL pattern. The problem we face however is that to obtain this result we need to restrict inputs, an action that runs counter to the Richness of the Base (Prince and Smolensky 1993/2004), a basic tenet of OT that states that no prohibitions should be imposed on inputs. This forces us to consider an input that is not moraically specified.

(2) CL in OT --- non-moraic input

	/CV <sup>μ</sup> C/	NoCODA	MAX-μ	MAX-SEG	DEP-μ
a.	CV <sup>μ</sup>			*	
b.	CV <sup>μμ</sup>			*	*!

This time, MAX-μ is no longer applicable since both outputs retain all input moras. MAX-SEG is equally satisfied by both candidates, therefore the decision is passed onto

DEP- $\mu$ . This chooses the candidate that lacks lengthening (2a), because it avoids mora insertion. As a matter of fact, given that the violations of (2a) are a subset of those incurred by (2b), (2a) harmonically bounds (2b), which means that any ranking of the constraints in (2) would favour the non-lengthened output.

Consequently, we run into a dilemma where both options are equally unappealing; we either need to impose input moraicity, but this violates ROTB (1), or we conform to ROTB, but then half of the time, we produce the incorrect results (2).

Numerous OT analyses have been devised to account for CL, but as I will show later on, all require some retreat from the basic premises of the OT model as sketched above, since they either introduce intermediate levels (Sprouse 1997, McCarthy 2003c, McCarthy 2005) or require moraic specification (Topintzi 2005b). Others attempt to avoid these problems by making use of notions such as segmental faithfulness (Hermans 2001) or as preservation of the numerical integrity of segments (Lee 1996). A variant of the former approach treats CL as coalescence (Sumner 1999) and seems more promising. All of these however fail in different respects.

For this reason, I will provide an alternative model, which aspires to solve some of the problems other accounts face. It is a fully parallelistic account and makes no use of implicit [e.g. sympathy (McCarthy 2003c), enriched inputs (Sprouse 1997)] or explicit serialism [OT-CC (McCarthy 2005); Stratal OT (Kiparsky 2000), which - as far as I know - has not offered an account of CL]. Moreover, it makes no reference to underlying moraicity and does not require implementation of stipulatory mechanisms such as WBYP. These are two obvious advantages over previous accounts.

Additionally, it can account for two famously problematic instances of CL. One involves CL after onset loss (Samothraki Greek) and the other CL in a language that lacks pre-existing syllable weight contrast (Piro). Both are predicted not to occur under Hayes' account where CL is all about mora preservation. For Hayes (and for that matter in all other moraic models other than the current one), onsets never carry moras, thus their deletion cannot cause lengthening. Furthermore, it is claimed that CL can only occur in languages which already show a syllable weight contrast, thus in cases like Piro where no long vowels arise and there is no evidence that coda consonants are moraic, CL should be impossible.

Based on Samothraki Greek and Piro, I argue that these predictions are incorrect and construct a model where such cases are allowed to emerge. Their rarity - particularly CL after onset loss - is attributed to different reasons independent of moraicity.

### 5.3 CL as position preservation

The basic idea I will pursue is that CL is not about mora preservation, but about position preservation via a mora. An input segment needs to have an output correspondent. There are two ways it can achieve this. It may either survive in the output intact (or with some of its identity features altered) or it may delete, in which case CL occurs via the addition of a mora. CL is then just seen as an alternative strategy for a segment to show up in the output with lengthening acting as a ‘cue’ for the lost segment. This can be coded in the constraint termed Position Correspondence (POSCORR).

- (3) POSCORR: An input segment must have an output correspondent either segmentally by means of a root node or prosodically by means of a mora

To illustrate, let us consider a classic synchronic case of CL. I use the following hypothetical example where the root /kan/ may either surface on its own or be followed by vowel-initial (-a) or consonant-initial (-ta) suffixes<sup>1</sup>.

- (4)    a. /kan/            →    [ka:]  
       b. /kan-ta/        →    [ka:ta]  
       c. /kan-a/        →    [kana]

(4a-b) show that when the root-final consonant appears in syllable final position, it deletes. On this occasion, CL occurs, producing lengthening of the preceding vowel. In (4c), no deletion occurs, since the consonant /n/ is no longer in syllable-final position. At the same time, no lengthening occurs either. How are we to account for these facts in a parallelistic account? The next tableaux will shed some light on this issue.

To facilitate their reading, I will be using indices to identify corresponding positions. To avoid too many indices, moras of input segments will have no indices.

---

<sup>1</sup> Alternations like the ones in (4) indicate that the consonant that deletes cannot be underlyingly moraic. This can be seen if one compares (4a, b) with (4c). In the former case, /n/ may be considered moraic since its deletion leads to lengthening, but in (c) no deletion occurs, even though the /n/ is now in what is usually considered to be a non-moraic position - an onset. The absence of lengthening could either be attributed to the lack of input mora for /n/ or to the retention in /n/ of an input mora, but then /n/ in (c) would have to be considered a moraic onset. Since this is not ordinarily accepted, the mora would need to be somehow realized in a position other than the onset. An obvious candidate is the vowel before, so then we should get [ka:na], which of course does not happen. In addition, under standard moraic accounts such as Hayes (1989) or Morén (1999/2001), an underlyingly moraic consonant denotes a geminate, but as (4c) highlights, no geminate arises. All these observations lead to the conclusion that the consonant /n/ cannot underlyingly bear a mora.

Only the inserted mora [in bold] due to CL will have an index to mark the mapping to the deleted segment.

(5) /kan/ → [ka:]

/k <sub>1</sub> a <sub>2</sub> <sup>μ</sup> n <sub>3</sub> /	NOCODA	POSCORR	DEP-μ
a. ..a <sub>2</sub> <sup>μ</sup> n <sub>3</sub>	*!		
b. ..a <sub>2</sub> <sup>μ</sup> n <sub>3</sub> <sup>μ</sup>	*!		*
c. ..a <sub>2</sub> <sup>μ</sup>		*!	
d. ..a <sub>2</sub> <sup>μ<sub>3</sub></sup>			*

The first two candidates retain the coda consonant and thus violate NOCODA, even though the second also assigns a mora to this consonant. The third candidate deletes the coda, but fails to present any correspondent of /n/, either segmental or prosodic. As a result, POSCORR is violated. The winning candidate is therefore (d), because it manages to simultaneously satisfy NOCODA but also retain a correspondent of the position occupied by /n/ by lengthening the previous vowel (indicated by the added μ and the index 3 which identifies it with the deleted consonant). All this happens at the cost of DEP-μ only.

Of course, there is one more important candidate we should have considered, namely *ka*<sub>2,3</sub><sup>μ</sup>. This has no lengthening, but presents coalescence of the deleted segment with the previous vowel. While this does not seem very likely, given that in such case we should also probably get nasalization of the vowel, we should consider this competitor too. By ranking the constraint against coalescence (UNIFORMITY) [McCarthy and Prince 1995] highly enough, this candidate will be easily excluded as in (7a)<sup>2</sup>.

(6) UNIFORMITY: No element of the output has more than one correspondent in the input, i.e. no coalescence

(7) /kan/ → [ka:]

/k <sub>1</sub> a <sub>2</sub> <sup>μ</sup> n <sub>3</sub> /	UNIFORMITY	POSCORR	DEP-μ
a. ..a <sub>2,3</sub> <sup>μ</sup>	*!		
b. ..a <sub>2</sub> <sup>μ<sub>3</sub></sup>			*

<sup>2</sup> An additional alternative somehow analogous to (7a) would have mora fusion (although μ<sub>3</sub> does not necessarily have a correspondent in the input): [k<sub>1</sub>a<sub>2</sub><sup>μ<sub>2,3</sub></sup>], but this would apparently violate some sort of UNIFORMITY with respect to moras.

This same analysis should work for /kan-ta/ → [ka:ta]. Now, let us see whether this ranking also works in /kan-a/ → [kana].

(8) /kan-a/ → [kana]<sup>3</sup>

	/k <sub>1</sub> a <sub>2</sub> <sup>μ</sup> n <sub>3</sub> -a <sub>4</sub> <sup>μ</sup> /	UNIFORMITY	*n/CODA	POSCORR	DEP-μ
a.	..a <sub>2</sub> <sup>μ</sup> n <sub>3</sub> a <sub>4</sub> <sup>μ</sup>				
b.	..a <sub>2</sub> <sup>μμ</sup> n <sub>3</sub> a <sub>4</sub> <sup>μ</sup>				*!
c.	..a <sub>2</sub> <sup>μ</sup> a <sub>4</sub> <sup>μ</sup>			*!	

This time there's no trigger for /n/-deletion (8c), so there's no reason why the input segment /n<sub>3</sub>/ should not have a segmental correspondent (8a). A prosodic correspondent by means of an inserted mora would satisfy POSCORR but only superfluously at the expense of DEP-μ (8b). Thus, we can see that the right results obtain. More generally, the pattern for CL involves the following:

(9) *Pattern for CL*

- a trigger for deletion of segments in particular syllabic positions must exist (markedness constraint M)
- POSCORR >> DEP-μ ensures that CL occurs, i.e. insertion of a mora identified with the position of the deleted input segment
- UNIFORMITY >> DEP-μ: prosodic rather than segmental identification is preferred

This system can also easily generate the absence of CL, as shown in (10). A hypothetical example of this sort also appears in /kan/ → [ka] illustrated in (11).

(10) *Deletion of a segment not followed by CL* --- preliminary version:

- requires minimal re-ranking: DEP-μ >> POSCORR implying that no trace (segmental or prosodic) of the deleted segment is left behind

<sup>3</sup> This raises the question of what the winning candidate in a language without CL would look like. This is tackled next. We could also consider a candidate like: [..a<sub>2</sub><sup>μμ</sup>a<sub>4</sub><sup>μ</sup>], that would have the violations of (8b), plus a violation of either ONSET or, if the vocalic sequence was wholly syllabified in a single syllable, \*3μ.

(11) *Lack of CL: /kan/ → [ka]*

	/k <sub>1</sub> a <sub>2</sub> <sup>μ</sup> n <sub>3</sub> /	UNIFORMITY	*n/CODA	DEP-μ	POSCORR
a.	..a <sub>2</sub> <sup>μ</sup> n <sub>3</sub>		*!		
b.	..a <sub>2</sub> <sup>μ</sup> n <sub>3</sub> <sup>μ</sup>		*!	*	
c.	..a <sub>2</sub> <sup>μ</sup>				*
d.	..a <sub>2</sub> <sup>μμ<sub>3</sub></sup>			*!	
e.	..a <sub>2,3</sub> <sup>μ</sup>	*!			

This is a good result, because obvious alternative analyses of CL face several problems. In an analysis à la Hayes, serial derivation has to be utilized under which WBYP acts before the coda consonant is deleted so that this receives a mora. After coda deletion, lengthening occurs. The absence of CL is attributed to the inactivity of WBYP or to its application after coda deletion. This is only possible however if the process is seen derivationally. The other possibility which would avoid this problem involves underlying moraic specification. This produces several problems as we have seen in the discussion of (4).

Note however that (c) is not the only conceivable winner in (11). As hinted in (9), high-ranking POSCORR dictates the need for a position correspondent, but due to the ranking UNIFORMITY >> DEP-μ, the prosodic option through a mora is chosen. Thus, CL occurs. But there is another possibility. If all else stays the same, but there is a minimal re-ranking so that DEP-μ >> UNIFORMITY, there will still be a position correspondent, but this time it will take the form of a segmental one. This ranking would select (11/12e) instead as illustrated in the tableau in (12). In effect, this is another way to produce the lack of CL.

(12) *Lack of CL: /kan/ → [ka]*

	/k <sub>1</sub> a <sub>2</sub> <sup>μ</sup> n <sub>3</sub> /	*n/CODA	POSCORR	DEP-μ	UNIFORMITY
a.	..a <sub>2</sub> <sup>μ</sup> n <sub>3</sub>	*!			
b.	..a <sub>2</sub> <sup>μ</sup> n <sub>3</sub> <sup>μ</sup>	*!		*	
c.	..a <sub>2</sub> <sup>μ</sup>		*!		
d.	..a <sub>2</sub> <sup>μμ<sub>3</sub></sup>			*!	
e.	..a <sub>2,3</sub> <sup>μ</sup>				*

In most instances where CL fails to occur, it is difficult to decide between (11c) and (12e), so in principle, both options are available. There is however a reason why I am inclined to choose (11c) over (12e) whenever data do not suggest otherwise. (12e) involves fusion, which entails that it should be possible to recognise whether the

segment is the product of coalescence. For instance, in the case of /kan/ becoming [ka], we could argue that fusion is involved had the resulting vowel shown nasalization, i.e. as in [kã]. This is indeed the case in French, as shown below. Consonant deletion is followed by nasalization, which suggests that the winning candidate in French would be the product of the ranking in (12).

(13) *French: [Y] ~ [Vn] alternations in masculine-feminine* (Tranel 1987: 70-71)<sup>4</sup>

bon [bõ]	vs.	bonne [bõn]	‘good’
fin [fẽ]	vs.	fine [fin]	‘thin’
certain [sertẽ]	vs.	certaine [serten]	‘sure’

A similar example is found in Portuguese (Mateus and d’Andrade 2002), where it is argued that nasalized vowels are the product of an underlying sequence of a vowel and a nasal. For instance, the prefix /in-/ appears as [in-] before a vowel, but as a nasalized vowel [ĩ] before a consonant. In the latter case, presumably nasal consonant deletion has occurred, with the nasal feature being left behind and linked to the preceding vowel.

(14) *Portuguese nasalized vowels* (Mateus and d’Andrade 2002: 22)

a. V-initial stems				
<i>acabado</i>	‘finished’	<i>inacabado</i>	[inẽkẽbádu]	‘unfinished’
<i>oportuno</i>	‘opportune’	<i>inoportuno</i>	[inopurtúnu]	‘inopportune’
b. C-initial stems				
<i>capaz</i>	‘able’	<i>incapaz</i>	[ĩkẽpáf]	‘unable’
<i>posto</i>	‘put in place’	<i>imposto</i>	[ĩpóftu]	‘tax’

This kind of evidence however is not always available. Consequently, whenever such direct evidence is lacking, I will assume that the language prefers to delete the position altogether leaving no relic of it behind, i.e. as in (11). For concreteness, though I will update the pattern for lack of CL accordingly.

(15) *Deletion of a segment not followed by CL* --- final version:

- i) no trace of deleted position: M, UNIFORMITY, DEP-μ >> POSCORR, (cf. (11))
- ii) segmental correspondent: M, POSCORR, DEP-μ >> UNIFORMITY (cf. (12))

<sup>4</sup> Tranel (1987: 49-50) also discusses lengthening in French. There are certain conditions under which this may occur. As far as nasal vowels are concerned, lengthening applies when they are in a stressed and closed syllable. In the cases mentioned here, the nasal vowel is not in a closed syllable, therefore no lengthening may show up.



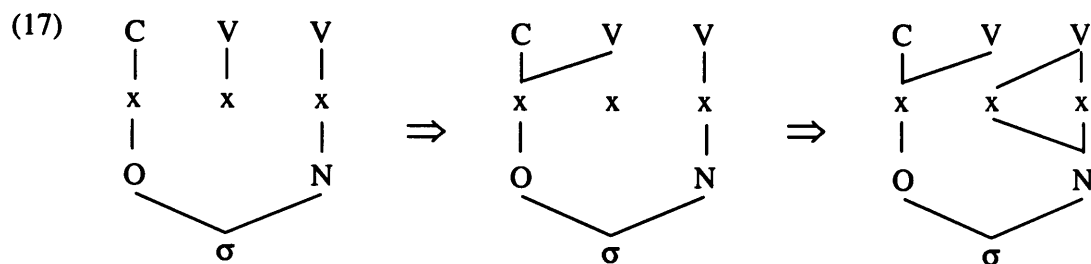
A patent question comes to mind. How about cases of CL where no segment is deleted? The obvious examples involve Bantu prenasalization and glide formation. In the former, Luganda inputs like /ba-ntu/ surface as [ba:<sup>n</sup>tu] ‘people’, while in the latter /mu-iko/ becomes [m<sup>w</sup>i:ko] ‘trowel’. No segment is lost here so why should CL occur in the first place? The answer lies in the definition of POSCORR repeated below.

- (16) POSCORR: An input segment must have an output correspondent either segmentally by means of a *root node* or prosodically by means of a mora

Work by Clements (1986), Sagey (1986) and Rosenthal (1994) suggests that the nasal or the high vowel in the sequences mentioned above has no correspondent in terms of a root node. In the output, the complex segment consists of a single root node which contains the nasal features or the ones of the high vowel. But since there is no segmental correspondent and assuming that POSCORR >> DEP-μ holds, some correspondent is required. This can only take the form of a prosodic correspondent by means of lengthening of the following vowel.

Kinyarwanda is a useful example for this claim (Sagey 1986). This language admits sequences of a consonant followed by a glide. Compelling evidence that this is a single complex segment with multiple articulations that consists of just one root node comes from the fact that Kinyarwanda does not allow complex onsets. More specifically, German loanwords that involve complex onsets are not preserved intact in Kinyarwanda (as predicted had the language allowed complex onsets), but are split by epenthesis so that simplex onsets are created, e.g. *Republik* > *repuburika*, *Präsident* > *perezida*, etc. Complex segments are allowed however, therefore they must be a single root node.

While I am not concerned with the specific features or the geometry involved in complex segments of this sort (see e.g. Sagey 1986, Lombardi 1990), the point which is of importance for us is that these segments are simply one root node. For instance, in Sagey (1986: 75), the representation of glide formation and subsequent CL is in (17) and corresponds to Kinyarwanda examples like the ones in (18):

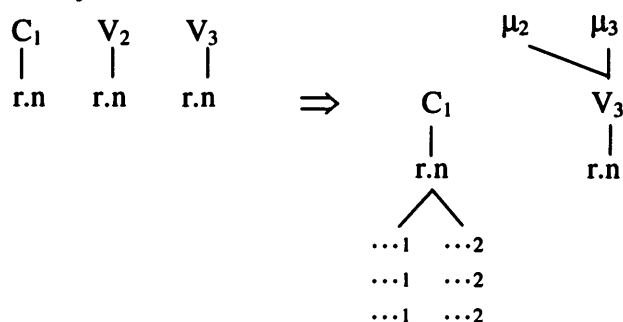


(18) /ku-i-Bon-a/ → [k<sup>w</sup>i:Bona]<sup>5</sup> ‘to see oneself’

/ku-gu-ir-a/ → [kug<sup>w</sup>i:ra] ‘to fall on’

The first vowel creates a complex segment with the preceding consonant under a single timing slot (or root node). The first vowel’s timing slot re-associates with the second vowel leading to CL. We could translate these representations to the current model in the following way:

(19) *Glide formation and CL*



An input sequence such as /C<sub>1</sub>V<sub>2</sub>V<sub>3</sub>/ then becomes [C<sub>1</sub>V<sub>3</sub><sup>μ<sub>2</sub>μ<sub>3</sub></sup>]. The features of V<sub>2</sub> are included in the complex segment C<sub>1</sub> which consists of one root node. V<sub>3</sub> on the other hand lengthens so that it provides a prosodic correspondent for V<sub>2</sub>, namely μ<sub>2</sub>. An identical process holds for examples that involve prenasalization<sup>6</sup>.

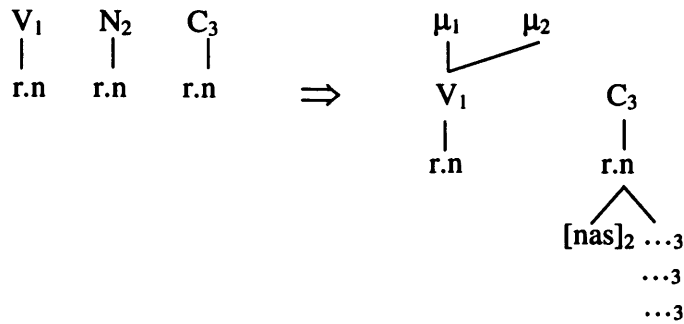
(20) /ba-nde/ → [ba:<sup>n</sup>de] ‘who?’

/ku-ngana/ → [ku:<sup>n</sup>gana] ‘to be equal’

<sup>5</sup> I choose here a transcription like [k<sup>w</sup>i:Bona] instead of Sagey’s [kwiiBona] so that I simultaneously stress the fact that we are talking about complex segments with multiple articulations as well as long vowels (rather than a sequence of identical vowels).

<sup>6</sup> The representation in (21) is in accordance with Sagey’s observation that “in Guarani, prenasalized stops are derived by a process of nasal harmony that spreads just the feature nasal. Thus, the resulting prenasalized stop must be branching just for the feature [nasal] and not for any class nodes” (1986: 50).

(21) *Prenasalization and CL*



An example of these facts using an OT tableau is shown immediately below<sup>7</sup>.

(22) /ba-nde/ → [ba:<sup>n</sup>de]

/b <sub>1</sub> a <sub>2</sub> <sup>μ</sup> n <sub>3</sub> d <sub>4</sub> e <sub>5</sub> <sup>μ</sup> /	UNIFORMITY	PRENASALIZATION	POS CORR	DEP-μ
a. b <sub>1</sub> a <sub>2</sub> <sup>μ</sup> . <sup>n</sup> d <sub>3,4</sub> e <sub>5</sub> <sup>μ</sup>	*!			
b. b <sub>1</sub> a <sub>2</sub> <sup>μ</sup> .n <sub>3</sub> d <sub>4</sub> e <sub>5</sub> <sup>μ</sup>		*!		*
c. b <sub>1</sub> a <sub>2</sub> <sup>μ</sup> . <sup>n<sub>3</sub></sup> d <sub>4</sub> e <sub>5</sub> <sup>μ</sup>			*!	
d. b <sub>1</sub> a <sub>2</sub> <sup>μ<sub>2</sub>μ<sub>3</sub></sup> . <sup>n</sup> d <sub>4</sub> e <sub>5</sub> <sup>μ</sup>				*

Candidate (b) fails to emerge with prenasalization, so it violates whichever constraint is responsible for this violation (informally called here PRENASALIZATION). The other candidates all emerge with prenasalization. The first one presents fusion, thus it violates UNIFORMITY. Among the two remaining rivals, (c) loses since its representation implies that the nasal consonant appears with its own root node. We have argued that there is empirical evidence suggesting that this is an incorrect representation of prenasalized stops, which consist of a single root node. Thus, in (c) there is no position correspondent for *n*<sub>3</sub>, therefore such form fails POSCORR. (d) wins since its only violation is the addition of a mora as an output prosodic correspondent for /n/.

## 5.4 Samothraki Greek

### 5.4.1 The data

With this much as background for the suggested CL model, we can proceed to the analysis of Samothraki Greek<sup>8</sup>. The reason Samothraki Greek (SamG) is interesting is

<sup>7</sup> The language bans codas, so I do not consider candidates that have codas.

because it presents data with /r/-deletion only from an onset position, but not from a coda one. Moreover, such deletion causes compensatory lengthening (CL) of the *following* vowel. This of course seems as a counterexample to more familiar cases of CL, where CL occurs after coda loss, but not after onset loss (Hayes 1989).

Unless indicated otherwise, data come from Katsanis (1996; henceforth K). Some extra data are due to some additional sources and from personal communication with Maria Tsolaki, a native speaker of Samothraki Greek and Marianna Ronga at the Aristotle University of Thessaloniki, who has knowledge of the data too<sup>9</sup>. Let us first present the /r/-facts: /r/ stays on two occasions: i) when it is word-final and ii) in sequences VrCV, i.e. when in coda position<sup>10</sup>.

(23) *Coda /r/ stays* [here and throughout all glosses are mine]

fanár	‘lantern’ (K: 48)	arpázu	‘I grab’ (K: 48)
figár	‘moon’ (K: 58)	karpós	‘seed’ (K: 48)

In all other instances, /r/ deletes. In the case of singleton onsets, /r/ deletion leads to vowel lengthening, but only word-initially (24). Word-medially no such lengthening occurs (25)<sup>11</sup>.

(24) *Deletion of /r/ initially and lengthening* (K: 50-51)

ra > a:	rafts > á:fts	‘tailor (masc.)’
ri > i:	riɣan > í:ɣan	‘oregano’
ru > u:	rúxa > ú:xa	‘clothes’
re > e:	réma > é:ma	‘stream’
ro > o:	róɣa > ó:ɣa	‘nipple, berry (of a grape)’

<sup>8</sup> This analysis improves significantly on the one proposed in Topintzi (2005b). There, underlying moraity of /r/ was an important assumption for the analysis. This is no longer required. Moreover, it was predicted that word-medial /r/-deletion leads to a slight phonetic lengthening which is not really perceived. On the contrary, now we can straightforwardly express the fact that no lengthening occurs in this environment at all. Some other points have been altered too. For comparison, the reader is referred to that work.

<sup>9</sup> SamG shares with other Northern Greek dialects the raising of stressless *e*, *o* to *i*, *u* respectively, e.g. *péde* > *pédi* ‘five’, *potamós* > *putamós* ‘river’ and the loss of underlying *i* and *u* (with some exceptions) *tiyáni* > *tyan* ‘frying pan’, *kufós* > *kfos* ‘deaf’. These are tangential to our current focus so I abstract away from them.

<sup>10</sup> /r/-deletion but no lengthening is also reported in the dialect of Taifiri of Kallipolis and occasionally in the dialect of Kapi in Lesvos (Papadopoulos 1927). He also mentions /l/-deletion in Samothraki Greek, a non-productive and practically no more attested process also acknowledged in Katsanis (1996) and Afroudakis (p.c. 29/12/05). Unlike them however, Papadopoulos claims that sometimes /l/-deletion leads to CL. Thus, we get interesting examples where both /r/- and /l/-deletion apply with subsequent CL, e.g. *lutrá* > *u:tá*: ‘baths’.

<sup>11</sup> Andriotis (1926-1928) suggests that /r/-deletion and lengthening occurs in a coda position too, e.g. *várka* > *váka* ‘boat’. Other works of the same period, e.g. Papadopoulos (1927) agree with more recent accounts such as Krekoukias (1964), Afroudakis (1985) and Katsanis (1996), who report that coda /r/ survives.

(25) *Deletion of /r/ word-medially and no lengthening* (K: 52)

aro > ao	θaró > θaó	'I reckon'
iru > iu	léftirus > léftius	'free'
are > ae	varéʎ > vaéʎ	'barrel'
iri > ii	θiríða > θíiða	'pigeon-hole'
eru > eu	kséru > kséu	'I know'
uri > ui	lurí > luí	'strap, strip'
era > ia	méra > mía	'day'
ara > aa	skára > skáa	'grill'
ure > ue	kurévo > kuévu	'I cut someone's hair'

In complex onset clusters, /r/ again deletes, but lengthening occurs in all positions. Numerous examples are given in K: 54-55, 59. Some are presented here:

(26) */r/ in onset cluster+V+C: deletion and lengthening*

a) biconsonantal clusters

pr+o > po:	prótos > pó:tus	'first'
vr+i > vi:	vrisi > ví:s'	'tap'
fr+e > fe:	fréna > fé:na	'brakes'
xr+o > xo:	xróma > xó:ma	'colour'
kr+a > ka:	krató > ka:tó	'I hold'
yr+a > ya:	yráfo > yá:fu	'I write'
θr+o > θo:	θrónos > θó:nus	'throne'
ðr+a > ða:	ðrákos > ðá:kus	'dragon'
br+e > be:	yabré > yabé:	'bridegroom'
dr+u > du:	ðédro > ðédu:	'tree'
tr+u > tu:	metrún > mitú:n	'they count'

b) triconsonantal clusters

spr+a > spa:	áspra > áspa:	'white'
xtr+a > xta:	éxtra > éxta:	'hostility'
zdr+u > zdu:	sidrufçá > zdu:fçá	'company (of people)'
ftr+a > fta:	ráftra > á:fta:	'tailor (fem.)'

Things are however different when the Cr+V sequence is followed by another vowel, i.e. Cr+V+V. The output of the sequence  $C + r + i/e + V$  is not  $C + i:/e: + V$ ; rather it is  $C + i + rjV$  without r-deletion or lengthening, but with metathesis and glide formation. As it stands, the analysis presented below also suggests that if the first vowel is other than *i/e* - namely one of the three remaining vowels *a, o, u* - the same kind of output should occur. Two points are worthy of discussion here. First, I have been unable to find any data that would fit the desired profile, i.e.  $C + r + a/o/u + V$ . Modern Greek words

like *fráula*, *krúo*, *práos*, *akróasi* are not used in the dialect (Ronga p.c. 9/3/05), thus we cannot yet test what happens in  $Cr+V_{[a,o,u]}+V$  sequences. Moreover, while empirical confirmation is absent, we do not actually expect to get metathesis and glide formation for these cases for independent reasons. Given that the vowels involved are back, a back glide *w* would be required, but this is missing from the dialect, as is missing from Standard Modern Greek too. With this caveat in mind, I will build an analysis designed to only make reference to  $C + r + i/e + V$  sequences.

(27)  $C + r + i/e + V \rightarrow C + i + rjV$  Ronga (p.c.)

priakóni > pirjakón	“jagged file used to sharpen knives”
alétria > alétirja	“plough (plural)”
tría > tirjá	“three”

Katsanis describes exactly the same phenomenon but with reference to velars only, i.e.  $Velar + r + i/e + V \rightarrow Velar + \hat{i} + rjV$ . An additional process applies here, namely centralization of front vowels *i/e* to  $\hat{i}$  (or  $\hat{e}$  occasionally)<sup>12</sup>.

(28)  $Velar + r + i/e + V \rightarrow Velar + \hat{i} + rjV$  [all from K: 71]

áyrios > áyîrjus	‘wild’	axríastos > axîrjastus	‘unneeded’
kréas > kîrjás	‘meat’	yriá > yîrjá	‘old woman’

Centralization of front vowels in the environment  $k/g/x/\gamma + r + i/e$  generally happens, i.e. even when a consonant follows (compare (29) with (26))<sup>13</sup>.

(29)  $Velar + r + i/e + C \rightarrow Velar + \hat{i}:\hat{e}: + C$

gr+i > gî:	grízos > gî:zus	‘gray’	(K: 56)
kr+i > kî:	krínu > kî:nu	‘I judge’	(K: 56)
kr+e > kê:	kremnós > kê:mnús	‘precipice’	(K: 72)

It should thus be evident that while vowel centralization is clearly related to velar consonants, the absence of /r/-deletion and the emergence of a glide in (27) and (28) is independent of the quality of the consonant; it relates to the presence of a  $V_{front}+V$ . In

<sup>12</sup> As the exact phonetics of  $\hat{i} / \hat{e}$  is quite unclear to me, I maintain Katsanis’ representation in this respect. However according to Ronga (p.c. 9/3/05),  $[\hat{i}]$  is most likely IPA  $[i]$  and  $[\hat{e}]$  is  $[\epsilon]$ . She maintains that  $[\hat{i}]$  is the unstressed realization of  $[\hat{e}]$ .

<sup>13</sup> According to Maria Tsolaki, some of the words above are instead pronounced as: *axêrjastus*, *kirjás*’ (where the final *s* is palatalized) and *gê:mnus*.

what follows, I will abstract away from centralization and focus on the /r/ loss and lengthening facts.

#### 5.4.1.1 Summary of the patterns

The preceding patterns can be summarized in the following table:

#### (30) SamG /r/ patterns

		Deletion	Lengthening	Other
Coda /r/		NO	NO	
Singleton	word-initially	YES	YES	
onset /r/	word-medially	YES	NO	
Complex onset /r/	Cr+V+C	YES	YES	(also when C=velar & V=front→centralization)
	Cr+V <sub>[i,e]</sub> +V	NO	NO	glide appearance +metathesis (+centralization as above)

### 5.4.2 The analysis

#### 5.4.2.1 On the placelessness of SamG /r/ and constraints required


One assumption that is going to prove crucial for the subsequent discussion is the claim that the SamG /r/ is placeless. Such placelessness will serve to explain not only why /r/ deletes from onset position in the first place, but will also provide the basis for a specific case of vowel spreading (cf. section §5.4.2.6).

The placelessness of /r/ receives support from other languages where a similar claim has also been put forward such as Yoruba (Akinlabi 1993) or English (Rice 1992). For instance, Rice observes that onset clusters \*tl \*dl \*pw and \*bw are impossible in English. Using evidence from similar cases like English, she claims that (onset) consonants cannot be identical in terms of Place structure. But then this generates a problem; /tr/ and /dr/ onset sequences are allowed. If both /r/ and /l/ are coronal, then why are r-coronal clusters accepted but l-coronal clusters are bad? A possible answer is that /r/ is placeless. As a result, coronal+r sequences are expected since they differ in terms of place of articulation.


The implication of /r/'s placelessness in Samothraki Greek is that /r/ will be able to survive in codas where placelessness is often accepted or required [cf. Selayarese ? (Rice 1992), Japanese ŋ (Yip 1991)], but not in onsets, because placeful onsets are preferred. Technically this point can be implemented by means of positional

markedness (e.g. Zoll 1998). The positional markedness schema includes a context-specific markedness constraint that dominates a faithfulness constraint and this in its turn dominates the context-free markedness constraint as in:  $M_{\text{Positional}} \gg F \gg M$ . The relevant constraints here are:  $*\text{ONS/PLACELESS} \gg \text{MAX-SEG} \gg * \text{PLACELESS}$ .

(31)

	/rar/	*ONS/PLACELESS	MAX-SEG	*PLACELESS
a.	rar	*!		**
b.	ra	*!	*	*
c.	a		**!	
 d.	ar		*	*

(32)

	/sas/	*ONS/PLACELESS	MAX-SEG	*PLACELESS
 a.	sas			
b.	sa		*!	
c.	a		**!	
d.	as		*!	

For an input like /rar/, only [ar] will manage to survive, as it exhibits minimal deletion and gets rid of onset /r/ thus avoiding a  $*\text{ONS/PLACELESS}$  violation. On the contrary, input /sas/ will emerge faithfully since there is no trigger for any MAX violations.

To simplify things however, I will collectively present the results attained by using the following constraint.

(33)  $*\text{ONSET/r: /r/ is disallowed in onset position}$

Our analysis will also be making use of the following constraints, most of which we have already seen in §5.3, while some more will be added as we move along.

(34) UNIFORMITY: No coalescence

POSCORR: An input segment must have an output correspondent either segmentally by means of a root node or prosodically by means of a mora

DEP- $\mu$ : Do not insert moras in the output

LINEARITY: No metathesis



#### 5.4.2.2 Coda /r/

Let us first start with the easy case. In codas, /r/ does not delete (23), thus any lengthening is gratuitous (35a-b). While no ranking argument can be provided by the coda data, by looking ahead to the onset analysis, the ranking UNIFORMITY, \*ONSET/r, POSCORR >> DEP- $\mu$  is proposed. /r/ of course does not delete since being in a coda protects it from a \*ONSET/r violation (35a).

(35) no CL when /r/ in coda: UNIFORMITY, \*ONSET/r, POSCORR >> DEP- $\mu$

	/k <sub>1</sub> a <sub>2</sub> <sup><math>\mu</math></sup> r <sub>3</sub> p <sub>4</sub> o <sub>5</sub> <sup><math>\mu</math></sup> s <sub>6</sub> /	UNIFORMITY	*ONSET/r	POSCORR	DEP- $\mu$
☞ a.	..a <sub>2</sub> <sup><math>\mu</math></sup> r <sub>3</sub> ..				
b.	..a <sub>2</sub> <sup><math>\mu</math></sup> r <sub>3</sub> <sup><math>\mu</math></sup> ..				*!
c.	..a <sub>2</sub> <sup><math>\mu\mu</math></sup> ..				*!

#### 5.4.2.3 Singleton /r/ word-initially

Word-initially, /r/ deletion is forced by \*ONSET/r, which now becomes active (24). The ranking UNIFORMITY, \*ONSET/r, POSCORR >> DEP- $\mu$  is justified. The first three candidates are wiped out as all violate one of the top-ranked constraints by presenting coalescence, a placeless onset or by failing to show up with a correspondent for /r<sub>1</sub>/. The final candidate manages to pass all these constraints, only at the expense of low-ranked DEP- $\mu$ . As a result, the initial vowel lengthens.

(36) CL of singleton /r/ word-initially<sup>14</sup>: UNIFORM, \*ONS/r, POSCORR >> DEP- $\mu$

	/r <sub>1</sub> a <sub>2</sub> <sup><math>\mu</math></sup> s <sub>3</sub> u <sub>4</sub> <sup><math>\mu</math></sup> /	UNIFORMITY	*ONSET/r	POSCORR	DEP- $\mu$
a.	a <sub>1,2</sub> <sup><math>\mu</math></sup> ..	*!			
b.	r <sub>1</sub> a <sub>2</sub> <sup><math>\mu</math></sup> ..		*!		
c.	a <sub>2</sub> <sup><math>\mu</math></sup> ..			*!	
☞ d.	a <sub>2</sub> <sup><math>\mu\mu</math></sup> ..				*

<sup>14</sup> A candidate which would lengthen the consonant instead, i.e. as.sa at the very least violates \*GEM, which seems to be undominated in SamG, since the dialect possesses no geminates. Moreover, such a candidate may also present problems in terms of locality, i.e. as in a<sub>2</sub> <sup>$\mu$</sup> s<sub>3</sub> <sup>$\mu$</sup> .sa<sub>4</sub>, an issue I set aside here. However, the constraint against long vowels NLV (Rosenthal 1994) must be low-ranked to allow lengthening.

#### 5.4.2.4 Singleton /r/ word-medially

In contrast to the word-initial position, word-medially no lengthening occurs (25). While this may seem at first glance puzzling, it proves quite a sensible thing. Imagine what would happen to an input like  $/l_1u_2^\mu r_3i_4^\mu/$  if lengthening took place. We would either expect *lu:i*  $[l_1u_2^{\mu 2\mu 3}i_4^\mu]$  or *lui:*  $[l_1u_2^\mu i_4^{\mu 3\mu 4}]$ . In other words, a sequence V:V (or VVV) would be created. As it seems, such sequences tend to be avoided in languages as observed by Kavitskaya (2002: 43) who claims that the ban on V:V “... *can be motivated by perceptual properties of VV sequences: since vowel-to-vowel transitions (emphasis added mine) are always very long, a two-vowel sequence is not likely to be re-interpreted as a three-vowel one*”. I would like to propose that this could be encoded in the markedness constraint in (37), which of course is very highly-ranked in SamG.

(37) \*SUPER-LONG VOCALIC HIATUS (\*S-L VH) / \*V:V

There is a reason why I have changed Kavitskaya’s term from ‘transitions’ to ‘hiatus’ in the markedness constraint. As Moira Yip (p.c.) points out, we cannot talk about ‘transitions’ in cases where the vowels flanking *r* are the same. Keeping the term ‘transitions’ would entail on the one hand correctly that when the surrounding vowels are different, no lengthening occurs after *r*-deletion, e.g. *luri* > *luɪ*, \**lu:i*, but on the other hand incorrectly that in a word like *θiriða* we would get *θi:iða*, instead of the actual \**θiðiða*, that lacks lengthening (cf. (25)). For this reason, the more neutral term ‘hiatus’ is chosen that also bans \**θi:iða*.

(38) no CL in singleton /r/ word-medially: UNIFORMITY, \*S-L VH, \*ONSET/r >> POSCORR >> DEP-μ

	$/l_1u_2^\mu r_3i_4^\mu/$	*ONSET/r	*S-L VH	POS CORR	DEP-μ
a.	$..u_2^\mu r_3i_4^\mu..$	*!			
b.	$..u_2^{\mu 3}i_4^\mu..$		*!		*
c.	$..u_2^\mu i_4^{\mu 3}..$		*!		*
d.	$..u_2^\mu i_4^\mu..$			*	

\*S-L VH ensures that (b) *lu:i* and (c) *lui:* are eliminated. (a) fails, because it does not delete the onset /r/. This time the winning candidate is actually one where POSCORR cannot be satisfied, as in (d). No correspondent - segmental or prosodic - for /r<sub>3</sub>/ exists,

thus the winner shows no lengthening. As a result, our ranking needs to be slightly modified. The relationship between (a) and (d) offers the ranking argument \*ONSET/r >> POSCORR, yielding the adapted ranking:

(39) SamG CL (to be revised): UNIFORMITY, \*S-L VH, \*ONS/r >> POSCORR >> DEP-μ

#### 5.4.2.5 Complex clusters of the type Cr+V+C - the simple(r) case

We have now dealt with simplex /r/. It remains to see what happens with /r/ in complex clusters. I will first consider the case where the cluster is followed by a V+C sequence ((26) and (29)). As we know, the output of a cluster of this type involves /r/ deletion and lengthening as in (d). Given the grammar employed so far, the right results obtain in a manner by now familiar.

(40) CL of /r/ in Cr+V+C<sup>15</sup>: UNIFORMITY, \*ONS/r >> POSCORR >> DEP-μ

/m <sub>1</sub> i <sub>2</sub> <sup>μ</sup> t <sub>3</sub> r <sub>4</sub> u <sub>5</sub> <sup>μ</sup> n <sub>6</sub> /	UNIFORMITY	*ONSET/r	POSCORR	DEP-μ
a. ...t <sub>3</sub> u <sub>4,5</sub> <sup>μ</sup> n <sub>6</sub>	*!			
b. ...t <sub>3</sub> r <sub>4</sub> u <sub>5</sub> <sup>μ</sup> n <sub>6</sub>		*!		
c. ...t <sub>3</sub> u <sub>5</sub> <sup>μ</sup> n <sub>6</sub>			*!	
d. ...t <sub>3</sub> u <sub>5</sub> <sup>μμ</sup> n <sub>6</sub>				*

#### 5.4.2.6 Complex clusters of the type Cr+i/e+V - the difficult case

In cases where the cluster is followed by i/e + V, no /r/ deletion occurs, but also no lengthening takes place ((27) and (28)). The question posed then is why for an input like /ayrius/ we get [áyîrjus] and not \*[áyî:jus]? Katsanis considers two approaches that involve derivational epenthesis and deletion and both of which he ends up finding problematic (K: 57). I present some additional reasons why these are implausible.

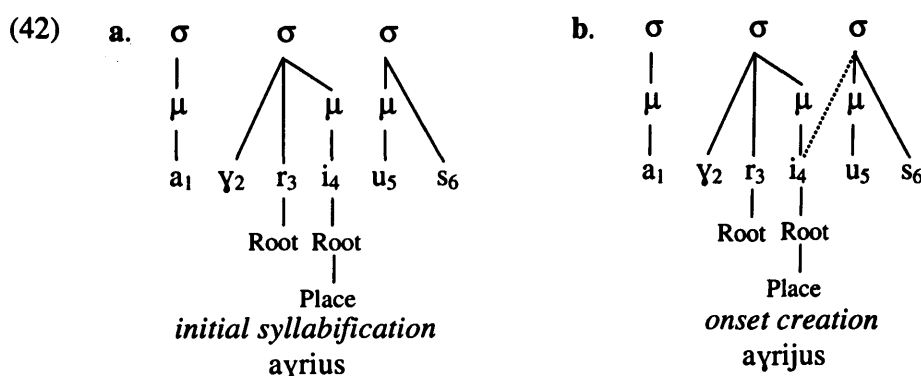
(41) i) áyrius → áyrijus (j-epenthesis) → áyîijus (r-deletion) → áyîjus (coalescence)  
→ áyîrjus (r-anaptyxis/epenthesis)

<sup>15</sup> A candidate like m<sub>1</sub><sup>μ</sup>t<sub>3</sub><sup>μ</sup>ru<sup>μ</sup>n<sub>6</sub>, i.e. [m<sub>1</sub>i<sub>2</sub>t<sub>3</sub><sup>μμ</sup>u<sub>5</sub><sup>μ</sup>n<sub>6</sub>] is excluded because it would entail that the onset is moraic. Assuming that \*MORAIC ONSET is high-ranked, this competitor is avoided. It is interesting however to note that while the current proposal allows CL because of onset loss, this does not have to relate to moraic onsets. As we will see later on, the claim is that any segment loss can result in CL, provided certain conditions are met.

ii) *áyrius* → *áyrijus* (j-epenthesis) → *áyirijus* (i-epenthesis between yr) → *áyîrjus* (second-i-deletion)

The problems (41i) faces are the following. First, /r/-epenthesis seems unlikely as it is a process generally unprecedented in Greek dialectology. But even if it was grounded, it is odd why /r/ should delete only to re-emerge later in the derivation. Moreover, why should the high central vowel coalesce, given that lengthening is not only allowed, but is in fact necessary in Cr+i/e+C? (41ii) is similarly troublesome. No good trigger for /i/-epenthesis exists, because SamG permits complex onset clusters e.g. *klévu* (K: 63), *kmar* (K: 64), *zmar* (K: 67)) and resolves complex onset clusters with /r/ by deleting it. So why should the language choose /i/-epenthesis instead? In addition, it is extremely bizarre to argue that the second /i/ deletes, as no markedness pressure seems to be applicable here.

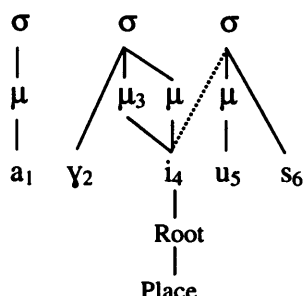
The alternative I offer is very different from both these approaches; I propose that what really goes on is *metathesis*, i.e. *r+i/e becomes i/e+r*. As a result, */r/ syllabifies in a coda and survives without getting deleted*. Finally, due to */r/’s placelessness* (cf. §5.4.2.1), *i/e can spread rightward over the /r/ and form an onset for the following syllable*. Recall from §5.4.1 that the current analysis in principle extends to cases where  $V_1$  is *a*, *o* or *u*, but apart from lacking empirical confirmation, glide formation is not expected, since it would require the back glide *w* which is missing from the language. The proposal should become clearer if we consider the input */ayrius/* and the possible representations its output could take stepwise [I will abstract away from the centralization facts]. Consider first (42).



(42a) is what I call ‘initial syllabification’, i.e. the structure we should expect if nothing at all have happened. (42b) is the structure we would expect after onset creation. The

next logical step would then be to get *ayi:jus* (43) by /r/-deletion and subsequent lengthening, but this is not what happens. Why should this configuration be suboptimal?

(43) What *does not* happen - multiple linking in *ayi:jus*



Observe that in this structure, the segment /i/ has three links to prosodic constituents (two to moras, one to a syllable). I would like to suggest that this is a configuration avoided in languages, similar in spirit with other restrictions against ternary structures, such as \*3μ. Thus, (43) can be excluded by the following proposed constraint:

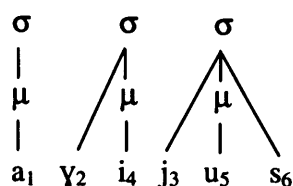
(44) \*3 LINKS: No ternary branching originating from a single segment

Such constraint on the surface seems to overlap with \*S-L VH presented in (37) that bans V:V sequences. However, while \*3 LINKS can eliminate (43), the same cannot be said for \*S-L VH. The reason is that the intervening onset glide can be considered to ruin the environment upon which \*S-L VH is relevant. Nonetheless, \*S-L VH can be applicable if one considers other candidates like *ayi:us* or *ayiu:s*, i.e. in cases where no glide appears in an onset position<sup>16</sup>. Only \*SL-VH but not \*3 LINKS could deal with rejecting these candidates.

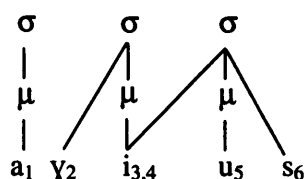
But there are more candidates to consider. The correspondent for /r<sub>3</sub>/ could for instance be one that changes features of /r/ and turns it to a glide. This is not a particularly plausible candidate, but even if possible, IDENT-[F] would easily rule it out (45a). Another, slightly more plausible candidate, involves fusion, but as we have seen throughout the analysis, UNIFORMITY violations deal with this too (45b).

<sup>16</sup> We could alternatively provide an onset by inserting a glide instead of spreading. This would avoid the \*S-L VH violation, but would violate DEP-SEG, the constraint against segment insertion, which I claim is highly ranked in the language.

(45) a. IDENT violation

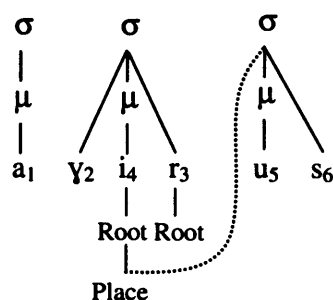


b. UNIFORMITY violation



This leaves us with the following representation.

(46) What *does* happen



This is based on /r/'s placelessness which permits /i/ to spread its place features over /r/ and onto the onset of the next syllable. At the same time, by having a metathesis between /i/ and /r/, /r/ now appears in a coda position, where it can survive (cf. §5.4.2.2) and thus no deletion occurs. Since /r/ has a segmental correspondent of itself, it satisfies POSCORR and therefore no lengthening takes place. While this form is consistent with \*ONSET/r, UNIFORMITY, POSCORR, \*S-L VH, \*3 LINKS, IDENT-[F] (as well as \*MORAIC ONSET and DEP-SEG discussed in footnotes), it presents violations of two constraints that had not been previously discussed. One is LINEARITY which bans metathesis, as it happens here since [...r<sub>3</sub>i<sub>4</sub>...] → [...i<sub>4</sub>r<sub>3</sub>...]. The other is INTEGRITY which penalises splitting of a segment into two, as it happens here with i<sub>4</sub>.

To see how the right results are obtained through our tableaux, let us first consider some candidates we can easily exclude as they violate high-ranked constraints.

(47) No CL but instead metathesis: less interesting candidates

	/a <sub>1</sub> <sup>μ</sup> γ <sub>2</sub> r <sub>3</sub> i <sub>4</sub> <sup>μ</sup> u <sub>5</sub> <sup>μ</sup> s <sub>6</sub> /	*3 LINKS	*S-L VH	*ONSET/r	POS CORR	DEP-μ
(42) a.	..r <sub>3</sub> i <sub>4</sub> <sup>μ</sup> u <sub>5</sub> <sup>μ</sup> ..			*!		
b.	..i <sub>4</sub> <sup>μ<sub>3</sub>μ<sub>4</sub></sup> u <sub>5</sub> <sup>μ<sub>3</sub></sup> ..		*!			*!
c.	..i <sub>4</sub> <sup>μ<sub>4</sub></sup> u <sub>5</sub> <sup>μ<sub>3</sub>μ<sub>5</sub></sup> ..		*!			*!
(43) d.	..i <sub>4</sub> <sup>μ<sub>3</sub>μ<sub>4</sub></sup> j <sub>4</sub> u <sub>5</sub> <sup>μ<sub>5</sub></sup> ..	*!				

The remaining candidates can be accounted for by this updated ranking:

(48) IDENT, UNIFORMITY >> INTEGRITY, LINEARITY >> DEP- $\mu$

(49) No CL but instead metathesis: more interesting candidates

	/a <sub>1</sub> <sup><math>\mu</math></sup> Y <sub>2</sub> r <sub>3</sub> i <sub>4</sub> <sup><math>\mu</math></sup> u <sub>5</sub> <sup><math>\mu</math></sup> s <sub>6</sub> /	UNIFORM -ITY	IDENT	POS CORR	INTEGR -ITY	LINEAR -ITY	DEP- $\mu$
(45a)	a. ..i <sub>4</sub> <sup><math>\mu</math></sup> j <sub>3</sub> u <sub>5</sub> <sup><math>\mu</math></sup> ..		*!			*	
(45b)	b. ..i <sub>3,4</sub> <sup><math>\mu</math></sup> u <sub>5</sub> <sup><math>\mu</math></sup> ..	*!					
(46)	c. ..i <sub>4</sub> <sup><math>\mu</math></sup> r <sub>3</sub> j <sub>4</sub> u <sub>5</sub> <sup><math>\mu</math></sup> ..				*	*	

INTEGRITY and LINEARITY have to be low-ranked, because the winning candidate violates both. Moreover, we can also form an important ranking argument that LINEARITY >> DEP- $\mu$  as shown by re-examining simple data from CL in complex onsets.

(50) LINEARITY >> DEP- $\mu$

	/k <sub>1</sub> r <sub>2</sub> a <sub>3</sub> <sup><math>\mu</math></sup> t <sub>4</sub> o <sub>5</sub> <sup><math>\mu</math></sup> /	LINEARITY	DEP- $\mu$
a.	..a <sub>3</sub> <sup><math>\mu</math></sup> r <sub>2</sub> ..	*!	
b.	..a <sub>3</sub> <sup><math>\mu_2\mu_3</math></sup> ..		*

This reveals that metathesis is not the preferred resolution for POSCORR satisfaction. Had it been, then we would expect metathesis rather than lengthening in cases involving Cr+V+C clusters. With all this in mind, the SamG grammar for r-deletion and CL involves:

(51) UNIFORMITY, \*3 LINKS, \*S-L VH, \*ONSET/r, IDENT >> POSCORR, INTEGRITY, LINEARITY >> DEP- $\mu$

As is evident, SamG CL presents the proposed pattern that CL needs to exhibit, namely: MARKEDNESS, UNIFORMITY >> POSCORR >> DEP- $\mu$ .

## 5.5 Alternatives to CL as position preservation

In this section, I explore alternatives to CL as position preservation with special reference to SamG. Note that reference to SamG is important, because it is one of the toughest cases to account for. Not only is it a case of CL with whatever problems this generally carries over for OT models, but also a case of onset CL. The latter implies that

even if general CL can be generally accounted for, as in e.g. Hayes (1989), the SamG data demonstrate additional difficulties owing to the special nature of the phenomenon.

### 5.5.1 Hayes (1989)

Let us then first see what Hayes (1989) has to say about the SamG data. Since in his model CL is attributed to mora preservation and since CL from onset loss is prohibited, he needs to offer a different account for this set of data. His attention is drawn on /CrV/ inputs which are claimed to have undergone V-epenthesis to split the cluster resulting in [CVrV]. Subsequently, intervocalic /r/ loss applied and merger of [VV] to [V:] followed.

Hayes bases this proposal on a suggestion by Newton (1972) that /i/ is an originally epenthetic vowel as in the development of *xartí* to *xaití* “paper” or *karpós* to *kaipós* “seed”. This would be analysed as [VrC] → [VriC] → [ViC]. Katsanis (1996: 48) mentions that these examples are reported in Heisenberg (1921), whose credibility is unquestionable. Nonetheless, Katsanis claims that such forms are not accepted by modern speakers. He eventually accepts Newton’s interpretation of the facts, but remains somewhat sceptic regarding its validity given that /i/ epenthesis is common in Imvros or Lesvos but does not emerge in Samothraki, as Newton himself observes with the exception of two examples, i.e. *tun-i-mikró* (MG: ton mikró) “the little one” and *babázim* (MG: babás mu) “my dad” (K: 51).

More importantly, apart from the complex onset data, Hayes also investigates the loss of word-medial /r/ and suggests that under his proposal, the absence of lengthening there is anticipated since there is no motivation for vowel epenthesis in a VrV sequence. Crucially, there is no discussion of the word-initial data, where the deletion of the singleton /r/ is actually followed by vowel lengthening. In analogy to the complex onset proposal, one would expect vowel epenthesis as a possible solution here, but no motivation exists for V-insertion in this case, since no cluster is involved. In fact, there should not be any reason for /r/-deletion either here, since in Hayes’ view, /r/-deletion only occurs intervocalically. Hayes’ analysis then fails to account for at least one pattern, that of word-initial /r/-loss and subsequent lengthening, therefore his solution is unsatisfactory.



### 5.5.2 OT accounts

While Hayes' analysis fails on empirical grounds, other more recent accounts largely fail on theoretical grounds. Three main lines of investigation will be reviewed here. The first implicitly or explicitly accepts that an intermediate level needs to be invoked for CL treatment (Sprouse 1997, McCarthy 2003c, McCarthy 2005). The second sees CL as segmental faithfulness (Hermans 2001) or as preservation of the numerical integrity of segments (Lee 1996) and the latter is a variant of the former approach which treats CL as coalescence (Sumner 1999). We will examine these in turn and end up saying that although Sumner's approach renders itself as the more promising one, some problems remain unresolved. For this reason, the model currently advocated is more successful.

#### 5.5.2.1 *Intermediate levels*

The first set of analyses reviewed here makes some use of the idea that intermediate representations or levels have to be introduced to account for CL and opacity more generally. Sprouse (1997) notes that only by using an intermediate representation is it possible to express the fact that the segment that deletes and causes CL, corresponds to what would be syllabified in a coda. Hence, he explicitly admits such a level. However, in order that he restricts the proposed system so that no unlimited levels are permitted, he puts forward the idea of enriched inputs, i.e. inputs which exhibit syllabification and moraification. Not any form can be an enriched input as enriched inputs are subject to particular conditions (for more discussion, see Sprouse (1997, section 5.1)). Despite all the attempts to restrict this model, it is still quite powerful and crucially makes use of the notion of an intermediate level, which is one aspect of the model that we would like to get rid of.

McCarthy's sympathy theory (2003c) faces more or less the same problems. The use of an intermediate level here is more covert by means of a sympathetic candidate, which never really emerges, but interacts with other candidates and eventually determines the correct output. Sympathy theory has faced a significant amount of criticism for a number of reasons, including the extreme complexity of the system, the introduction of intermediate forms and learnability issues as well as its failure to produce the correct results in certain occasions (cf. Bermúdez-Otero 2001). For all these reasons, I do not consider it a viable solution to the CL problem.

Recently, McCarthy (2005) has presented another model of a similar type dubbed OT-CC where CC stands for what he calls candidate chains. This framework resembles in certain aspects the harmonic serialism (HS) model of Prince and Smolensky (1993/2004). Unlike harmonic parallelism which includes just a single pass from input to output, evaluation in harmonic serialism is repeated until it converges to the correct output. However, the ranking assumed each time is the same (unlike Stratal OT, where in each stratum a different ranking is imposed). Crucially, under HS all candidates considered in each pass differ from the assumed input in just one respect (a property that inspires the use of ‘candidate chains’), e.g. for an input like /kat/, candidates that are considered can only be: {kat (faithful), kati (insertion), kač (palatalization)}, but not {kači} since the latter simultaneously involves palatalization and insertion. The postulated ranking at this pass will choose one candidate as the winner. Let us assume that this is [kati], which now becomes the input for the next pass. This time {kači} is a possible candidate since it only presents palatalization. The imposed ranking in this case could be one which selects {kači} so no further evaluation takes place.

McCarthy makes use of the basic ideas of HS and builds a framework which allegedly accounts for opacity. It is outside the scope of the current work to examine this framework in detail, especially given that it is not yet fully worked out, however a couple of notes are in place. First, note that this framework explicitly makes reference to intermediate stages, which seems to be a full circle back to serial derivational approaches. It remains to see whether such a costly move is justified. Second, McCarthy does not consider CL opacity in this work, but it seems unclear how exactly OT-CC can treat it.

To exemplify, suppose that we have an input such as /CV<sub>μ</sub>C/ which in the output becomes [CVV] through coda deletion and subsequent CL. According to OT-CC (see McCarthy 2005: example (8)), chains are made up of candidates which are either faithful to the input or - if less faithful - more unmarked than it. This is what is called harmonic improvement. Suppose then that in our example, coda loss occurs due to high-ranking NOCODA. This then means that the first candidate chain we need to consider is {CV<sub>μ</sub>C, CV<sub>μ</sub>}. The former candidate is totally faithful, but violates NOCODA while the latter is less faithful, but less marked. The problem however here is that the actual winner is CV<sub>μμ</sub>. To achieve this we should perhaps first include the candidate {CV<sub>μ</sub>C<sub>μ</sub>} in our candidate chain, so that {CV<sub>μ</sub>C, CV<sub>μ</sub>C<sub>μ</sub>}. Notice though that at this stage, this candidate is less faithful, because it incurs a DEP-μ violation, but presents no obvious

improvement in markedness (other than WBYP satisfaction, but such a constraint is contentious<sup>17</sup>), thus it should not be a candidate included in the candidate chain. But without such a candidate it is very unclear how the desirable CV<sub>μμ</sub> can be generated.

This hypothetical situation has been implemented by means of an analysis closer to an HS approach, which as McCarthy observes, faces difficulties with opacity. It may just as well be the case that an analysis strictly along OT-CC's lines may work. Even if it does however, it should again make reference to an intermediate stage - already problematic - plus it would require the use of PREC(A, B) constraints, which stipulate that violation of constraint A must precede violation of B. Once again, this suspiciously resembles rule ordering, bringing us back closer to a serial account.

This approach then seems to explicitly deny the property that started off OT in the first place and led to its successful implementation, namely parallelism. By accepting OT-CC we run the risk of re-introducing serialism in the model and thus reverting to older theories with all the problems these faced. Moreover, if it is the case that opacity has to be handled by means of serial derivations, one wonders why such an elaborate and complex system such as the one in OT-CC has to be employed instead of the more familiar and straightforward rule-based accounts. Given that parallelism is one of the most significant tenets of OT, I contend that at least until further justification, OT-CC is not a plausible explanation for CL opacity and perhaps for opacity in general.

#### *5.5.2.2 Segmental faithfulness*

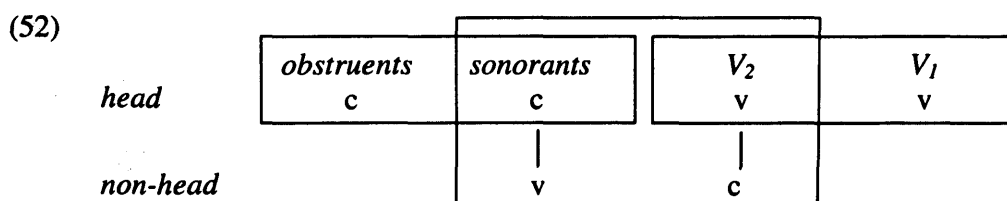
This type of approach shares with the coalescence approach (to be presented in the next section) the fact that it keeps track of the number of segments involved (indicated by means of indices). Other than that, some differences arise. Such analyses are akin to the current view, although the latter is not interested in the preservation of the number of segments, but in the preservation of the number of input positions, which can take the form of either segmental or prosodic elements.

---

<sup>17</sup> I mean 'contentious' here in a very specific way. Although WBYP, or its equivalent MORAIC CODA, are widely used in the literature (Morén 1999/2001; McGarrity 2003; Crosswhite 2001; Broselow, Chen and Huffman 1997 among many others), this constraint is basically a convenient stipulation to assign moraicity onto coda consonants, which would otherwise surface as non-moraic. Moreover, in many cases, WBYP is used alongside its opposite \*MORAIC CODA (for striking examples, it suffices just to look at Crosswhite 2001, or McGarrity's 2003 tableaux (24) and (25)). Both these constraints are markedness constraints, but it is unclear what the unmarked option for a language really is: should codas be moraic or not? I will not explore this issue in this work. It is nonetheless worthwhile noticing that while I make use of similar constraints too, I do not subscribe to any particular view of markedness. OT-CC on the other hand could only be viable if it assumed that moraic codas are the unmarked option for a language.

Lee (1996) attributes CL to the preservation of the “numerical integrity of segments of a morpheme”<sup>18</sup>. In this way, a form such as /C<sub>1</sub>V<sub>2</sub>C<sub>3</sub>C<sub>4</sub>V<sub>5</sub>/ becomes [C<sub>1</sub>V<sub>2,3</sub>C<sub>4</sub>V<sub>5</sub>]. This approach fails in many respects. First, there is no explanation why an output such as [C<sub>1</sub>V<sub>2,3</sub>C<sub>4</sub>V<sub>5</sub>], i.e. without lengthening, should be excluded. Lee claims that the proposed analysis can handle lots of CL instances without use of moras, but it seems that only the deletion of a mora could dispose of [C<sub>1</sub>V<sub>2,3</sub>C<sub>4</sub>V<sub>5</sub>]. What is more, as the analysis progresses, moras are explicitly used to account for CL cases such as Bantu prenasalization /muntu/ → [mu:<sup>n</sup>tu]. There is no point then to introduce an additional mechanism that keeps track of the segmental numerical integrity, if moras are still going to be used. Furthermore, underlying moras (cf. p. 21-26) and input syllabification (cf. p. 4, 22-24) play an important role in the generation of the patterns. Since input specification proves critical in this model, we again face the ROTB problems described in the beginning of this chapter.

A somewhat similar account appears in Hermans (2001), who also sees CL as segment preservation, but forms a more elaborate analysis so that he captures Hayes’ claim that there is no CL after onset loss. Hermans claims that segments in an input-output correspondence relation should be identical in terms of sonority. Technically, the major segmental classes are characterised by c- and v- elements (c- meaning low degree of periodic energy, and v- meaning high degree of periodic energy). Each segment has an obligatory c- or v-head, and, depending on its type, may also have a non-head element too. This yields the following categories:



V<sub>1</sub> refers to the first part of a long vowel, with V<sub>2</sub> being the second part of it. Obstruents only have a c-head element and V<sub>1</sub>’s have a v-head only. Sonorants are consonants with some vocalic feature (non-head), while V<sub>2</sub>’s are the reverse, by virtue of the fact that in some languages the second part of a long vowel is reduced<sup>19</sup>. The idea is that three natural classes are formed: i) those with a c-head, i.e. obstruents and sonorants, ii) those

<sup>18</sup> As far as I can see, there is no definition of the notion ‘numerical integrity’ anywhere, but we are being told how this can be achieved, i.e. by the general ranking MARKEDNESS >> MAX >> FAITH.

<sup>19</sup> I am not convinced that this renders the ‘v-head, c-non-head’ as V<sub>2</sub>’s, but for the sake of clarity, I am following Hermans’ suggestion.

with a v-head, i.e.  $V_1$  and  $V_2$  and iii) those whose makeup is as a whole identical, although the internal structure may differ, i.e. sonorants and  $V_2$ .

The idea is then that since segments need to belong to the same sonority class based on one of the categories above, effectively CL cannot occur after onset loss, because the product of that type of lengthening would occupy the  $V_1$  position. But  $V_1$  may only be in correspondence with segments of the  $V_2$  type, therefore such CL is impossible. On the other hand, CL after sonorant coda loss is possible, because the product of lengthening corresponds to the  $V_2$  position, and, as we have seen, sonorants and  $V_2$ 's may stand in correspondence.

There are two problems with this approach. The first one is that I contend that in light of the SamG data, it is wrong to conclude that CL from onset loss is impossible. Obviously, it occurs and while Hermans exceptionally seems to allow a special case of onset CL, i.e. that of  $/C_1V_2C_3V_4/ \rightarrow [C_1V_2V_3.V_4]$ , both SamG patterns of word-initial and complex onset loss are missed. This is an obvious defect.

Moreover, it seems to me that Hermans also misses another relatively standard pattern, namely that of obstruent coda loss and lengthening, e.g.  $/t_1a_2p_3/ \rightarrow [t_1a_2a_3]$ . Given that the only natural class obstruents are in is the one that also includes the sonorant consonants, I cannot see how an obstruent may delete and cause lengthening reflected on the  $V_2$ , since the latter belongs to a different class.

All in all, while his proposal is an interesting one and certainly insightful in some respects, it empirically suffers, thus it is not a promising alternative.

### 5.5.2.3 CL as coalescence (Sumner 1999)

This approach also focuses on segmental preservation through the use of numerical indices. There is a certain attractive feature that this analysis makes use of which makes it stand out as more promising than the other alternatives. More specifically, it does not refer to input moraicity. In fact, although moras are included in representations, CL is not seen as the product of mora preservation. Instead, what drives CL is the following ranking:

(53) BIPOSITION >> MAX-SEG

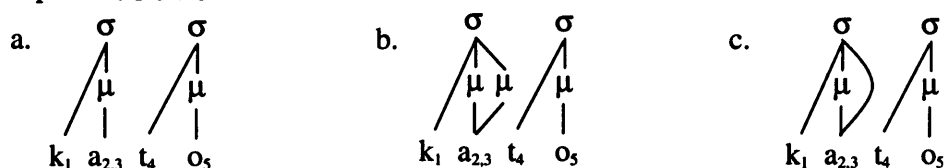
The former constraint is defined in the following way.

- (54) **BIPOSITION:** An output segment representing two input segments (denoted by subscripts) must be linked to two prosodic positions [Sumner 1999: 538]

What is handy in this constraint is the reference to ‘two prosodic positions’, without necessarily imposing linking to two moras. Let us see how this approach fares when considering the Samothraki Greek cases.

Things should become clearer once few possible representations are depicted. I shall first consider a rather straightforward case, such as /krató/ becoming [ka:tó]. Importantly, observe that I will consider a non-moraic input and see how the proposed analysis handles the facts.

- (55) Input: /k<sub>1</sub>r<sub>2</sub>a<sub>3</sub>t<sub>4</sub>o<sub>5</sub>/



The first representation fails **BIPOSITION** since the coalesced vowel only links to one prosodic position. The third representation avoids such a problem, since the vowel has two links, one to a mora and one to a syllable; nonetheless it exhibits a structure which is at the very least controversial. Bear in mind that Rubach (1998) argues for exactly such a structure in his discussion of Slovak glides. Curtis (2003) comments on some problems this representation carries with it. I will not discuss these here, but it suffices to mention that even if a representation like (c) is acceptable for high vowels as glides, it seems a highly unlikely one for other vowels. We are thus left with (b), a candidate which satisfies **BIPOSITION** by virtue of the extra mora the coalesced vowel is linked to. Under this approach then, CL indeed has to do with segmental preservation, a by-product of which is vowel or consonant lengthening. Let us see these results in a tableau:

(56) \*ONSET/r, BIPOSITION >> MAX-SEG

	/k <sub>1</sub> r <sub>2</sub> a <sub>3</sub> t <sub>4</sub> o <sub>5</sub> /	*ONSET/r	BIPOSITION	MAX-SEG
a.	k <sub>1</sub> r <sub>2</sub> a <sub>3</sub> t <sub>4</sub> o <sub>5</sub>	*!		
(55a) b.	$\begin{array}{c} \mu \\   \\ k_1 a_{2,3} t_4 o_5 \end{array}$		*!	
c.	$\begin{array}{c} \mu \\   \\ k_1 a_3 t_4 o_5 \end{array}$			*!
(55b) d.	$\begin{array}{c} \mu \quad \mu \\ \diagdown \quad / \\ k_1 a_{2,3} t_4 o_5 \end{array}$			

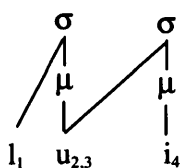
We have already discussed the candidates above with the exception of (56c). This is the candidate that presents no coalescence, so effectively /r/ has deleted altogether.

The advantage of this approach is that it needs make no reference to underlying moraicity or particular syllabic positions. What is more, it seems much more lenient than Hayes' model in admitting CL from onset loss, which seems particularly suitable for cases like Samothraki Greek. Nonetheless, Sumner acknowledges that there are cases where CL does not occur, since the segment deletes altogether and thus no lengthening applies. She is a bit cryptic at this point though in saying that there are independent constraints which decide whether segments delete or coalesce. Note that such constraints must at least dominate MAX-SEG and penalize (d) so that a candidate like (c) is preferred.

Assuming that such constraints are indeed plausible, then this approach produces good results. In particular, in Samothraki Greek, a reanalysis along Sumner's lines is possible. However, there is a single environment where her approach makes incorrect predictions. As a matter of fact, this problem also arises in her own example (23b), but she devotes no discussion on it.

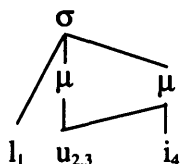
Consider our example /luri/ from (38). There we proposed a structure where no lengthening occurs, because if it did, an undesirable V:V structure would be generated. Note however that there is a structure that adheres to Sumner's proposal.

(57) /l<sub>1</sub>u<sub>2</sub>r<sub>3</sub>i<sub>4</sub>/ --- [l<sub>1</sub>u<sub>2,3</sub>.i<sub>4</sub>] = l<sub>1</sub>u<sub>2</sub>w<sub>3</sub>i<sub>4</sub>



Although this representation captures the lack of vowel lengthening after /r/-loss word-medially, it has the implication that the high back vowel acts as an onset of the following syllable, i.e. effectively something like *lu.wi*. The same problem arises even if take  $[l_1 u_2 i_{3,4}]$  as the output, i.e. *lu.ji*. These are clearly wrong since the actual output is *lui*. Perhaps, a remedy for this would be the following structure:

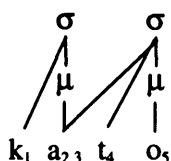
(58)  $/l_1 u_2 r_3 i_4/ \rightarrow [l_1 u_{2,3} i_4]$



The vowels here are represented as a diphthong, but it is quite unclear what this structure entails. Furthermore, it is equally possible that the index of /r/ is linked to the vowel *i*, but as it stands we cannot simply tell. Phonetically the prediction seems to be that *u* should incorporate about two thirds of the diphthong's duration, leaving only one third for *i*.

Finally, setting aside this specific case, it is not immediately obvious why under Sumner's approach the following structure should not be generally preferred.

(59)



This is equivalent to the problematic representation in (57), but is nonetheless appealing compared to the winner (56d), since it satisfies all constraints that (d) does, but also does not incur any DEP- $\mu$  violation. It is unclear how (59) could be excluded, other than making use of phonotactic considerations, especially given that Sumner wants to keep this configuration available since she makes explicit use of it on occasion. Indeed this remains a significant problematic aspect of this model which the present approach does not share.

### 5.5.3 A phonetically-based approach - Kavitskaya (2002)

Kavitskaya examines the SamG data too, and unlike Hayes (1989), she considers a fuller range of data, although not the complete set. In particular, there is no reference to the */ayrius/* →  $[\acute{a}y\acute{i}rj\acute{u}s]$  type of cases (cf. (27)-(28)).



On the whole, she proposes a model where CL is seen not as weight preservation, but as phonologization of inherent phonetic length. For instance, in CVC syllables, vowels are phonetically longer when followed by certain consonants whose transitions can be misheard as part of the vowel (i.e. sonorants, approximants). When these consonants delete, the “excess” length of vowels is now justified by phonologising it. Thus vowels are reinterpreted by listeners as phonemically longer. Since her model does not bear on the idea of CL as weight preservation, Kavitskaya is free to assert that onset deletion can cause CL when the consonants involved fit a certain profile.

In particular, she claims that *r* is vocalic enough to be reinterpreted as additional vowel length, as is the case in SamG. The fact that no lengthening occurs when *r* deletes in a word-medial position is attributed to a ban against VVV sequences, a position I have adopted as well. Incidentally, while such assertion also explains why no lengthening occurs in *Cr+V+V* sequences such as \**ayirus*, one can only guess why \**áyîrjus* is chosen instead. As a matter of fact, this pattern is not discussed at all. Moreover, the data referring to coda *r* come from Newton (1972), who claims that when *r* is in a coda position, then *i* is inserted between *r* and the following consonant. Subsequently *r* deletes as it is found intervocalically, e.g. *xartí* > *xarítí* > *xaití* ‘paper’. As we have seen earlier, in his recent study of the dialect, Katsanis (1996) claims that such forms are not accepted by modern speakers and questions the application of *i*-epenthesis generally in the dialect.

One way or another, nowadays there are words such as *karpós* or *arpázu* where the *r* is preserved intact syllable-finally (cf. (23)), while both Kavitskaya and Katsanis accept that *r* word-finally is preserved, e.g. *fanár*. Now the question is the following: if Kavitskaya is right in the reinterpretation of *r*’s length, then why does this fail to apply when *r* is in a coda position, as in the two cases mentioned just above? This is particularly disconcerting, since coda *r*-deletion and lengthening would be the prototypical situation rather than onset *r*-deletion and lengthening. Of course, with several amendments, such as the placelessness of *r*, much of the analysis can be rescued, but this presupposes a much more prominent role for phonology than what seems to be the author’s intention.

There is one additional question; if *r* is vocalic enough and reinterpreted as extra vowel length, then why no similar result obtains with more vocalic glides? While the status of glides in Modern Greek is notoriously controversial and confusing, there seems to be some evidence that there are some underlying glides as minimal pairs like *áðia* ‘permit, leave’ vs. *áðja* ‘empty-NEUTER-PL’ suggest. It would make sense to expect a

similar behaviour from glides as with the rhotic *r*, but there is no evidence for anything like that.

Finally, and on a more technical note, Kavitskaya (2002: 180-184) considers in passing a few analyses of CL within an OT framework, using the idea of moraic conservation. Importantly, she assumes that there is full syllabification and moraification in the input, a point which as we have shown is not only incompatible with core OT, but also generates wrong results. Thus, while Kavitskaya's approach has certain appealing properties and rightly permits CL from onset loss, her analysis of at least SamG exhibits some defects and less worked out features.

## 5.6 An extension: CL in Piro

It was previously mentioned that CL as position preservation has another significant advantage, which all other analyses cannot handle, unless certain modifications apply. Given that CL is not tied with moraic segments in certain syllabic positions, it is now expected that CL can occur as a result of deletion of segments that are unsyllabified or non-moraic. Piro CL can be viewed as an instance of this sort. Piro is an interesting example for an additional reason; it lacks phonemic long vowels and has no evidence for moraic codas, and yet undergoes CL rendering itself as a counter-argument to Hayes' (1989) claim that only languages with a pre-existing syllable weight contrast can participate in CL.

Piro is an Arawakan language spoken in Eastern Peru (Lin 1997, Matteson 1965). While there is consensus that all syllables in Piro are onsetful and open, some controversy exists over how the various consonant clusters are syllabified. Matteson (1965) argues for a (C)(C)CV structure, whereas Lin (1997) suggests that syllables are simply CV. Remaining consonants are unsyllabified. Lin observes that in all positions in the word, all possible clusters are allowed, ignoring sonority sequencing principles. Only some clusters are banned, since their cooccurrence leads to OCP violations by sharing similar place and/or manner of articulation.

### (60) *sample of Piro clusters*

tpa	'curve'	pto	'... 's group
mwenutu	'cheap'	wmahataya	'we lack'
wyoptota	'we receive'	ywalitxa	'hip'
ksu	'tube'	skota	'lower abdomen'
tmennu	'flaw'	mtenotu	'short'
smota	'blunt point'	msa	'empty corn cob'

The consonants which are not in a prevocalic position either surface as syllabic or followed by a very short epenthetic vowel, whose role is to facilitate consonant-consonant transitions. This vowel is inactive phonologically speaking, so it “is best treated as a phonetic phenomenon” (Lin 1997: 406).

Piro’s interesting property for our purposes is the fact that it has a ‘boundary vowel deletion’ rule (BVD) which deletes the final vowel of each lexical root and derived stem at each stage of suffixation. The process applies cyclically from the innermost domain to the outermost one. There are two exceptions to BVD. First, certain suffixes are lexically marked as not triggering the deletion process (see the example in (62)). BVD also fails to apply if it would lead to illicit or extra-long clusters (I am following Lin’s notation in marking the deleted vowel with the underscore and the non-deleted vowels with boldface).

(61) *Boundary vowel deletion (BVD)*

	nika + ya + waka + lu		‘to eat it there’
	to eat-LOC-place-it		
cycle 1	nika+ya	→	nik_ ya
cycle 2	nikya+waka	→	nikyawaka
cycle 3	nikyawaka+lu	→	nikyawak_lu

On the first cycle the stem vowel deletes, but this does not repeat on the next cycle, since deletion would lead to the illicit [kyw] cluster. On the final cycle BVD can again apply, because no similar risk arises. On other occasions, BVD is accompanied by subsequent consonant deletion which optionally or obligatorily leads to CL of the previous vowel (62). Since it is not my aim here to focus on the exact conditions under which BVD is blocked or CL applies (see Lin 1997, 1998), I will draw my attention onto the process itself as illustrated in the following example. For presentational purposes, I will assume that a simplified version of a constraint \*CCC is in effect banning three consonant clusters, whereas OCP forbids certain clusters that share place or manner of articulation. Both of these lead to consonant deletion and subsequent CL. Of course, a fully-fledged analysis of Piro requires consideration of all these cases.

(62) *BVD, C-deletion and CL* (non-BVD-triggering affixes are marked with capitals)

i) obligatory CL	
nika+ka → nikka → ni:ka	‘he is eaten’
hira+re-TA → hirreta → hi:reta	‘to drink’
xitxi+tši → xitxtši → xi:tši	‘foot’
ii) optional CL	

wane+YA+yi → waneyyi ~ wane:yi	'you have'
ruhi...TA+hima → ruhhimata ~ ru:himata	'it is said that he answers'

Significantly, Piro has no underlying weight contrasts. There are no geminates other than the ones which are the result of morphological concatenation. Moreover, the distinction between long and short vowels is not phonemic, but derived, since all long vowels are the outcome of CL. This fact, combined with the generalisation that all syllables in Piro are open pose a serious problem for standard moraic theory and CL.

As Lin correctly points out, according to Hayes (1989), CL can only occur in languages that have a pre-existing weight contrast. While Piro lacks such contrast, CL is still applicable. In addition, since consonants are not syllabified as codas, they are presumably syllabified as onsets. Again, Hayes prohibits CL from onset deletion. Piro therefore renders itself a challenge for moraic theory.

There are two basic assumptions that Lin makes in order to avoid these problems. First, she claims that these consonant clusters are not syllabified as complex onsets. Doing so - apart from the CL problem - makes it difficult to account for certain co-occurrence restrictions. For instance, if consonants of any type make legitimate onsets, it is difficult to explain why \*CCC should hold. Thus, Lin proposes that only the prevocalic consonant is syllabified in an onset. Other consonants before it remain unsyllabified. The absence of sonority restrictions and the possibility for virtually all combinations between consonants (other than the OCP-violating ones) is compatible with this proposal. In this way, Lin avoids the problem of CL after onset loss.

Second, these unsyllabified consonants are underlyingly moraic (cf. Bagemihl 1998)<sup>20</sup>. This move has two gains. First, it can account for the occurrence of CL; since the consonants are moraic, their deletion leads to lengthening due to high-ranking MAX-μ. Second, it allows Lin to modify Hayes' claim about the necessity of input syllable weight contrasts as a prerequisite for CL to the following generalisation: "a language that has the moraic and non-moraic contrast among consonants or the monomoraic and bimoraic contrast among vowels can exhibit CL" (Lin 1997: 424). According to Lin, Piro falls under the first category, since it has a contrast between unsyllabified moraic consonants and syllabified moraless onset consonants.

Notably, this distinction makes reference to output contrasts, which can be quite problematic and circular. In particular, imagine a language that not only lacks unsyllabified consonants altogether but also any positive evidence about the moraicity

---

<sup>20</sup> More accurately, her assumption is that all segments are underlyingly moraic. I will return to this issue later on.

of consonants, which could consequently be considered as non-moraic. The only way for this language to have CL is then by fulfilling the other part of the proposal, i.e. by making reference to the long vs. short vowel contrast. But given the proposed generalisation, it is actually possible that a language fits this criterion by presenting no other distinction but a V vs. VV contrast due to CL itself. This is of course circular, since the validity of CL is based on the moraicity contrast, which in its turn hinges upon the CL facts.

Nonetheless, Piro could actually be an example of such a language, if it is shown that unsyllabified consonants are not moraic. As I will argue later on, Piro unsyllabified consonants may just as well be non-moraic, thus Lin's generalisation fails. On the other hand, I will argue that since CL is not about mora preservation, input syllable weight or any modification of this sort is simply not applicable, which is why CL may even happen to languages like Piro with unsyllabified - moraic or not - consonants.

The biggest problem for Lin though is the input moraification she has to assume. While this does not complicate things in Piro, it runs into several problems once one considers most other languages. Technically, it also violates the Richness of the Base since it imposes restrictions to the input.

All these problems vanish in the alternative approach which sees CL as position preservation. The idea is the following; unsyllabified segments may lack syllable affiliation, but are like all other segments in requiring an output correspondent. As we have seen before, this may come in the guise of an entirely faithful output that preserves the segments in question or in the shape of moras which serve as the prosodic remnant of that position. When a consonant deletes, the first option is ruled out. The only solution then is CL. Notably, assignment of input moras plays no role in this. If we assume that the consonant that deletes lacks underlying moras, then we simply need to add a mora that conserves the lost position. This is possible because POSCORR >> DEP-μ. If the consonant on the other hand is assumed to include input moras, then that mora can act as the correspondent of the lost position. In this case DEP-μ is not violated at all. A sketch of this account is illustrated in the following example.

(63) /mtasa+xe/ → [mta:xe] --- first attempt

/m <sub>1</sub> t <sub>2</sub> a <sub>3</sub> s <sub>4</sub> a <sub>5</sub> +x <sub>6</sub> e <sub>7</sub> /	BVD	OCP <sub>f</sub>	POS CORR	DEP-μ
a. m <sub>1</sub> t <sub>2</sub> a <sub>3</sub> s <sub>4</sub> a <sub>5</sub> x <sub>6</sub> e <sub>7</sub>	*!			
b. m <sub>1</sub> t <sub>2</sub> a <sub>3</sub> s <sub>4</sub> x <sub>6</sub> e <sub>7</sub>		*!		
c. m <sub>1</sub> t <sub>2</sub> a <sub>3</sub> x <sub>6</sub> e <sub>7</sub>			*!	

 d.	$m_1 t_2 a_3^{\mu_3 \mu_4} x_6 e_7$				*
--	-------------------------------------	--	--	--	---

I follow Lin in utilizing the constraint  $OCP_f$  which militates against adjacent fricatives. This is violated in (63b) where /s/ and /x/ are found next to each other. The first candidate violates the constraint which forces the boundary vowel deletion process, informally stated here as BVD. Both these contenders are quickly ruled out. We are now left with (c) and (d). Both apply BVD and delete the fricative /s/ to avoid an  $OCP_f$  violation. The difference lies in the fact that (d) ensures an output correspondent for /s<sub>4</sub>/ by inserting a mora and thus lengthening the vowel /a<sub>3</sub>/. In this way it satisfies POSCORR. The DEP- $\mu$  violation is thus unimportant. (c) by failing to lengthen lacks such a correspondent and therefore loses with respect to POSCORR's evaluation.

Two issues are pertinent here. First, as we have said, nothing would change substantially had we assumed that the input consonants were moraic. While in (63), I deliberately make no claims about this point, it should be obvious that if /s/ also had a mora, then after the segment's deletion, no mora would need to be *added*; instead the pre-existing one would subsume the role of the segment's correspondent. Consequently, candidate (d) would again be the winner, but without any DEP- $\mu$  violation.

A more challenging question relates to the absence of lengthening after the vowel's deletion due to the application of BVD. More concretely, the winning candidate presents /a<sub>5</sub>/ deletion, but there seems to be no moraic correspondent for this segment. While it is quite unclear what the anticipated remedy should be in Piro, there is one possible way out of this. Perhaps, Piro provides evidence that POSCORR can be realised by means of its subparts POSCORR-C and POSCORR-V. The former requires a position correspondent for consonants and the latter for vowels. This should not come as a surprise since POSCORR's akin constraint MAX-SEG has often been claimed to require separation in its two variants: MAX-C and MAX-V (used in several analyses, e.g. Kager 1999 on Southeastern Tepehuan, Hall 2000 on Zoque, Kiparsky 2002 on Arabic, etc.).

For Piro then, all we would need to do is to rank POSCORR-C above DEP- $\mu$  so that its deletion is followed by subsequent lengthening, but POSCORR-V below DEP- $\mu$ , so that no lengthening is caused after vowel deletion. Of course one needs to note that under the assumption that vowels have input moras, /a<sub>5</sub> <sup>$\mu$</sup> / would already have a mora in the input which could be realised in the output, without violation of DEP- $\mu$ . Apparently this case would require some amendment against trimoraic syllables, so that the vowel does not become super long.

(64) /mtasa+xe/ → [mta:xe] --- revised version

/m <sub>1</sub> t <sub>2</sub> a <sub>3</sub> s <sub>4</sub> a <sub>5</sub> +x <sub>6</sub> e <sub>7</sub> /	BVD	OCP <sub>f</sub>	*3μ	POS CORR-C	DEP- μ	POS CORR-V
a. m <sub>1</sub> t <sub>2</sub> a <sub>3</sub> s <sub>4</sub> a <sub>5</sub> x <sub>6</sub> e <sub>7</sub>	*!					
b. m <sub>1</sub> t <sub>2</sub> a <sub>3</sub> s <sub>4</sub> x <sub>6</sub> e <sub>7</sub>		*!				*
c. m <sub>1</sub> t <sub>2</sub> a <sub>3</sub> <sup>μ<sub>3</sub>μ<sub>4</sub>μ<sub>5</sub></sup> x <sub>6</sub> e <sub>7</sub>			*!			
d. m <sub>1</sub> t <sub>2</sub> a <sub>3</sub> x <sub>6</sub> e <sub>7</sub>				*!		*
e. m <sub>1</sub> t <sub>2</sub> a <sub>3</sub> <sup>μ<sub>3</sub>μ<sub>4</sub></sup> x <sub>6</sub> e <sub>7</sub>					*	*

## 5.7 On the rarity of CL after onset loss

It should by now be evident that if CL works pretty much in the way described in this chapter, then reference to terms such as “CL after coda loss” or “CL after onset loss” do not make much sense, as there is no point in the ‘derivation’ during which these consonants syllabify. They are however descriptively convenient. Consequently, Hayes’ claim that CL occurs only after coda loss is significantly weakened. It makes much more sense to talk about CL after consonant deletion. Despite that, it is indeed true that there is a well-established asymmetry between coda and onset CL with the first being exceedingly common, and the latter increasingly rare.

There are certain reasons which provide some explanation for this discrepancy. First, it has to be noted that not all cases of segment deletion result in CL anyway, which of course entails that numerous cases of potential CL actually never really emerge. The Australian language Lardil serves as a good example (Kenstowicz 1994a and references cited therein)<sup>21</sup>. Data from Lardil stems reveal that while codas are permitted, these can only be coronals unless they are homorganic to the following onset in which case labials and velars are allowed too. Some examples are presented below.

(65) *Lardil stems and permitted codas*

- |      |             |         |                 |
|------|-------------|---------|-----------------|
| i)   | coronal     | wu.lun  | ‘fruit species’ |
|      |             | rel.ka  | ‘head’          |
|      |             | kar.mu  | ‘bone’          |
| ii)  | homorganic  | kuŋ.ka  | ‘groin’         |
| iii) | combination | ŋam.pit | ‘humpy’         |

<sup>21</sup> Another example is Diola-Fogny (Sapir 1965, Kager 1999), where deletion of a coda consonant in a heterorganic cluster occurs to avoid violation of the coda condition. No lengthening of the preceding vowel arises although phonemic length exists for all vowels (Sapir 1965: 6). For instance /let-ku-jaw/ surfaces as [lekujaw], \*le:kujaw ‘they won’t go’ or /jaw-bu-ŋar/ becomes [jabuŋar], \*ja:buŋar ‘voyager’.

Stems that show alternations between final noncoronals and Ø can be easily understood in the light of the generalisation above. Consonants that do not fit the profile above have to delete when in coda position (absolute form); on the other hand, when they can syllabify as onsets, they emerge intact (inflected form). Notably for our purposes, although a coda deletes, no lengthening occurs on the previous vowel, i.e. \**ɲalu:* or \**thurara:*.

(66) *Stem alternations: final C-deletion*

<u>absolute</u>	<u>inflected</u>	
ɲalu	ɲaluk-in	‘story’
thurara	thuraraŋ-in	‘shark’

In addition, Lardil also presents final vowel apocope in stems of three or more syllables. This time a vowel - which is unquestionably moraic - deletes, but again no CL applies.

(67) *Stem alternations: final V-apocope*

<u>absolute</u>	<u>inflected</u>	
yalul	yalulu-n	‘flame’
mayar	mayara-n	‘rainbow’

In fact, both V-apocope and final C-deletion may occur when the stem  $\geq 3\sigma$  and the consonant in question is labial or velar. Once more, CL does not apply.

(68) *Stem alternations: V-apocope & final C-deletion*

<u>absolute</u>	<u>inflected</u>	
putu	putuka-n	‘short’
tipiti	tipitipi-n	‘rock-cod species’

Hayes (1989) acknowledges that not all instances of coda deletion lead to CL. One easy way to explain this is that in those cases, WBYP has not applied and therefore there is no mora on the coda to compensate for after its deletion. Since the whole system ties moraicity and CL so closely, the prediction is that onsets should never trigger CL as they are non-moraic, while codas should trigger CL only when moraic. The problem now this account faces is that it ignores the effect of CL after nucleus loss. Since nuclei are always moraic<sup>22</sup>, their deletion should always cause CL. And yet as we have seen in Lardil this is not always the case.

<sup>22</sup> This is overwhelmingly the case. However, there are instances where some vowels have been claimed to contribute no mora. See for instance Shaw (2004) who argues that schwa in Mohawk is non-moraic.



Numerous instances of a similar kind can be easily traced in the literature. Consulting standard textbooks is a good start. In a quick search, I have located several examples where a vowel deletes and no lengthening occurs. Some of the examples include: Chukchee vowel apocope (Kenstowicz 1994a: 105), e.g. *nileq* ~ *nileqe-t* (absolute sg. ~ absolute pl.) ‘match’; Tangale vowel elision (Kenstowicz 1994a: 96); Klamath first stem (short) vowel syncope between a CV-prefix and a following CV syllable, e.g. *mac<sup>h</sup>a:t-ka* ‘listens’ ~ *sna-mc<sup>h</sup>a:t-i:la* ‘causes to hear’ (Odden 2005: 121), and Icelandic *i*-deletion and vowel syncope (Odden 2005: 189-190). Of course, in most cases things are more complicated, so the interested reader is invited to check these works and the references cited therein for details.

In the light of the discussion above, it should be evident that the connection between moraicity and CL is thus not as robust as previously claimed. Segment deletion, even if the segment is uncontroversially moraic, i.e. a vowel, is not guaranteed to be counterbalanced by CL. This fact makes the explanation that the absence of onset CL as the product of the lack of onset weight less convincing. Since the generalisation above holds equally for all syllable constituents, the absence of CL after onset loss may simply relate to the lack of CL occurring frequently in languages and not to the absence of onset weight per se. This then provides the first explanation of why no expectation for CL after onset loss is necessary. But this does not tell us much about the fact that onset CL specifically is not only fairly uncommon, but also indeed increasingly rare. I would like to suggest that there are some additional reasons at play which pertain specifically to onsets.

The most prominent is the fact that it is cross-linguistically the case that coda loss is overwhelmingly more common than onset loss. Onset loss in the beginning of the word is repeatedly observed in Australian languages, but in virtually no case has CL occurred. An important exception nonetheless seems to be Proto-Austronesian initial consonant loss, which according to Zewen (1977), has led to CL, relics of which can still be seen in Marshallese today (§7.2.2.2). The fact that onset loss is on its own quite rare obviously decreases the chance for CL to a large extent.

The other factor which seems to be at play is that deletion of an onset and subsequent lengthening can often lead to super-long vocalic hiatus, which as we have seen before, is universally dispreferred (Kavitskaya 2002). This type of hiatus would arise in virtually all cases where there is a singleton intervocalic onset, i.e. VCV, that gets deleted. By ranking the constraint \*S-L VH highly, we can ensure that lengthening of this type cannot occur. As a matter of fact, Samothraki Greek is an excellent example

of this type. /r/ deletes from an onset position, and causes lengthening whenever this would not violate \*S-L VH, namely word-initially and in a complex onset cluster. Word-medially, its deletion would result in the prohibited V:V configuration, thus no lengthening takes place. This makes the strong prediction that there will be no language with compensatory lengthening after onset loss word-medially, without equivalent lengthening word-initially.

More generally, since \*S-L VH seems to be cross-linguistically highly-ranked, it falls out that it frequently destroys the environment for CL after onset loss.

## 5.8 Conclusion

In this chapter, I have argued for an analysis that sees compensatory lengthening not as mora conservation, but instead as position preservation through a mora. A segment requires an output correspondent, either segmentally as a root node, or prosodically as a mora. This is dubbed POSCORR. However, when the segment deletes, only the prosodic correspondent solution remains. In the languages where POSCORR takes priority, lengthening through a mora occurs as a response to POSCORR satisfaction. The net effect is CL. In the languages where lengthening is banned, POSCORR is either sacrificed so that no output correspondent for the segment exists, or it is satisfied by having a segmental correspondent by means of fusion.

Moraicity is then seen as the remedy, rather than the trigger of CL. This analysis does not require reference to underlying moras (Lin 1997) or intermediate representations (Hayes 1989) to account for CL. Given a certain ranking, CL can occur in a familiar input-output manner, thus this solution is obviously advantageous because it does not compromise OT's orientation to the output. It also does not conflict with other important tenets, such as the Richness of the Base, which although not entirely uncontroversial, has remained influential.

But apart from being theoretically advantageous, this analysis has empirical merits too. Without any additional assumptions, it can account for the existence of CL in languages which present deletion of an onset or of unsyllabified material. The former is explicitly ruled out as an option in virtually all analyses of CL, while the latter is more controversial, but as Lin suggests, again it seems to be impossible in the standard moraic theory of Hayes (1989). If CL is about segment preservation through moras, then it is actually expected that we should get cases like that, because these too involve segments. Syllabification into an onset or no syllabification whatsoever is beside the

point. What matters is that there must be a prosodic correspondent for the deleted segment.

However note that as it stands, this analysis does not form a direct argument for onset moraicity, simply because it does not make any reference to retention of underlying moras. It treats all segments as equal and thus as equally able to cause CL<sup>23</sup>. The fact that CL from onset or unsyllabified-material loss is not as common, probably relates to the rarity of this kind of loss in the first place. Additionally in the former case, the usually high-ranking \*S-L VH can also block the process. This proposal then does not use CL as a testing ground over weight.

But bear in mind that the preceding analysis has been designed to conform to some of the major tenets of OT, such as avoidance of serialism (as serialism contradicts parallelism) and liberation from input restrictions (as input restriction contradicts ROTB). It is then essential to avoid making any reference to onset moraicity, and for that matter, to coda moraicity too. The important point, which this thesis aspires to have achieved, is to show that CL after onset loss exists and that it *can, but does not have to be*, the product of onset moraicity. In addition, as we will see in §7.2.2.3 when discussing data from Trique, there is supporting evidence that apart from CL which is caused by onset loss as in SamG, there is also *CL that causes onset lengthening* and produces a moraic onset.

In previous accounts that viewed CL as mora preservation, it was impossible to express SamG CL simply because such proposals denied onsets any possibility for weight contribution. This is not the case anymore. If one prefers to treat CL of the SamG type under the scope of more traditional accounts in terms of mora conservation, this thesis provides the tools to do so by permitting onset weight. As a matter of fact, an analysis of SamG making use of onset moraicity is already available (see Topintzi 2005b), but this nevertheless bears the - by now familiar - problems that all previous OT analyses of CL face.

The current analysis of course does not claim to be a full theory of CL, since this would require exhaustive investigation of the reported CL cases. This is by far beyond the scope of the present work. Presumably, certain additions and modifications would be required to account for the full inventory of CL cases. Then, both approaches, i.e. position preservation and mora conservation would need to be tested against them. No

---

<sup>23</sup> At a closer inspection, some refinement may be required, since it is an oversimplification to say that all segments are equal. For instance, epenthetic segments in some languages are invisible for certain processes (cf. for instance the discussion on Marshallese epenthetic yV- prefix in Hendricks (1999) and §6.4.2 here).

matter which proves to fare better, one thing is for sure. CL after onset loss is possible, either because every segment is eligible to cause CL (cf. Piro unsyllabified consonants) or because onsets are moraic, a conclusion compatible with the position advocated for in this thesis.

Having examined instances of coerced onset weight, I will now turn to distinctive onset weight which involves moraic geminates wholly syllabified in the onset. I will explore data from Pattani Malay, Trukese and Marshallese, and argue that onset moraic geminates are possible both word-initially and word-medially.

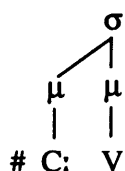
## Chapter 6

### Moraic onsets and geminates

#### 6.1 Introduction

In this chapter, I will be looking at languages with initial geminates which contribute to the moraicity of the syllable. I will claim that the introduction of the machinery of moraic onsets used in this thesis can provide an easy and straightforward account of initial moraic geminates. These are considered to be moraic onsets as shown in (1)<sup>1</sup>.

(1) *Initial moraic geminates as moraic onsets*



I will also show that this representation has been implicitly (Muller 1999) or explicitly (Hajek and Goedemans 2003) adopted in previous works and that it is advantageous over other representations such as Davis (1999) and Curtis (2003). I will then present data from Pattani Malay, Trukese and Marshallese and demonstrate how these support or are compatible with the structure in (1).

The chapter is structured as follows: In section 2, I evaluate the theoretical arguments for the ‘moracic onset’ analysis of initial moraic geminates. Section 3 deals with two languages which support such an account. One is Pattani Malay, where effects from initial geminates can be identified with reference to stress. I review Hajek’s and Goedemans’ (2003) analysis of the same facts, who also explicitly utilize moraic onsets. However, I point out several shortcomings of their proposal which can be eliminated by

---

<sup>1</sup> Throughout this chapter, I will include the length mark (:) in (1) and other similar representations. I am not of course claiming that length is represented on both the segmental (as part of the featural content of the segment) and timing tiers. I merely use it to remind the reader I am talking about geminates and not about singletons. This becomes quite important particularly since I claim that singleton moraic onsets receive the same representation as the geminate ones. Moreover, I will represent geminates as either C: or C<sub>i</sub>C<sub>i</sub>. The former is used in the discussion of Pattani Malay, the latter in the discussion of Trukese and Marshallese. This choice just follows the major sources of data for these languages and should facilitate comparison with them. On occasion, this rule may be overridden, if it serves analytical purposes. For instance, it is much easier to show the coda-onset syllabification of a geminate in C<sub>i</sub>C<sub>i</sub> instead of C:.

modifying certain aspects of their account. The other language is Trukese, whose gemination effect can be seen with respect to Minimality and final-mora deletion. The proposed analysis is based on work by Davis and Torretta (1998) and particularly Muller (1999). I show how a ‘moraic onset’ analysis is entirely compatible with the data at hand. In section 4, I return to an idea first advocated in section 2, namely that moraic onset geminates may also occur word-medially. I illustrate how this option is possible theoretically and exemplify it with data from Marshallese. In the final section, I briefly consider the relationship between geminate weight and length and also suggest a possible representation for long non-moraic onset consonants.

## 6.2 Theoretical background

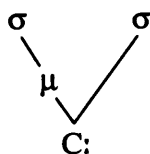
Many languages present a distinctive contrast between singleton and geminate consonants word-medially, e.g. Hungarian, Gujarati (Morén 1999/2001). Traditionally (cf. Hayes 1989), this distinction is held up to be one between underlyingly moraic consonants (geminates) versus non-moraic ones (singletons). This is shown in (2).

### (2) *Singletons vs. geminates in the input (Hayes 1989)*



According to this proposal, a word-medial geminate surfaces with a flopped structure where the first part of the geminate syllabifies in a coda (and carries a mora), and the second part directly associates to the onset of the following syllable. There are two advantages in this configuration. First, it captures the increased length of a geminate compared to a singleton. Second, the assignment of a mora to the coda expresses the fact that the first syllable of the two becomes heavy and as such can attract stress or participate in other weight sensitive phenomena.

### (3) *Flopped structure of geminates word-medially*

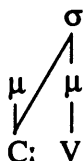


While this is the accepted representation of a geminate for many researchers, the proposal that geminates are cross-linguistically underlyingly moraic has been questioned

over the years, particularly in the light of languages with geminates, which nonetheless present no evidence about their moraicity, e.g. Malayalam (Tranel 1991), Cypriot Greek (Muller 2001, Arvaniti and Rose 2003) or Berawan and Tukang Besi (discussed in Chapter 7). Despite that, in the majority of languages, geminates are moraic, suggesting that apparent exceptions need to be re-analysed somehow (cf. Ham 2001 and §6.5.2).

A more interesting problem for our purposes is the existence of initial geminates<sup>2</sup>. While initial geminates are not as common as medial ones, several languages have them. In fact, there are certain languages which only allow initial geminates (Pattani Malay, Sa'ban and perhaps Nhaheun cited in Muller 2001 and Logbara and Oneida cited in Hajek and Goedemans 2003). The problem arises once we try to represent an initial geminate. If we attempt to extend the representation in (3) to initial positions, then again we end up with double linking, but this time, it is unclear where the left branch should attach to, since there is no coda available that can simultaneously host the mora and the first part of the geminate. Thus, Davis (1999) proposes (4) where the first part of the geminate remains unlinked.

(4) *Davis' (1999) representation of initial geminates*



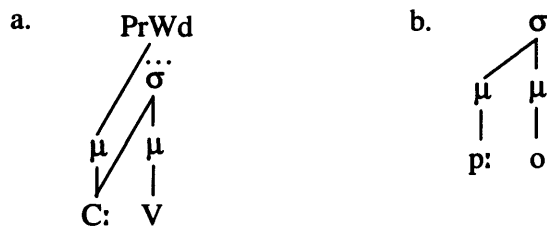
The obvious difficulty this proposal faces relates to its inability to count the first mora in weight processes, since a mora needs to be affiliated to higher prosodic structure, e.g. a foot or a PrWd (cf. Kiparsky 2002) in order to be computed. Here it links to neither, therefore, it will be impossible to add to the weight of a syllable<sup>3</sup>. This has further repercussions, because there are languages, such as the ones examined below, where the mora of an initial geminate matters for Word Minimality (Trukese) or stress (Pattani

<sup>2</sup> Word-final geminates are similarly problematic, but will not be tackled here. See Ham (2001) for a possible account.

<sup>3</sup> Hayes (1995) presents the moraic trochee as a possible foot type that several languages employ. Since this foot-type creates feet of two moras and as it is interested in moraic rather than syllable structure, it is conceivable that under a moraic trochee analysis, the configuration in (4) could constitute a bimoraic foot. However, moraic trochee analyses almost always assume that syllable integrity is nevertheless observed, so that moras belonging to the same syllable are grouped in the same foot rather than different ones. There are a few exceptions to this generalization such as Southern Paiute and Winnebago, both of which however are re-analysed in a way that syllable integrity is not violated. One analysis that explicitly challenges syllable integrity is that of Banawá (Everett, Buller and Buller 1993, Everett 1996, 1997). The case in (4) is however slightly different since it proposes to incorporate in the same foot syllabic and extra-syllabic material, rather than material which belongs to two different syllables.

Malay). Consequently, the structure in (4) has to be discarded<sup>4</sup>. As a response to this problem, Curtis (2003) proposes (5a), but also accepts that the representation in (5b) is structurally possible, although incompatible with standard moraic theory.

(5) *Possible representations of initial geminates according to Curtis*



Even though according to (5a), the consonantal mora can now be computed for weight considerations, this representation seems to partially identify initial geminates with unsyllabified consonants like the ones surfacing in Bella Coola (Bagemihl 1991), Piro (Lin 1997) or Arabic (Kiparsky 2002). The status of geminates seems to be different though and provides no evidence against the presence of syllabification. Consequently, (5a) has to be dispensed with too.

We are thus left with (5b), which, I claim, is actually the right representation of initial moraic geminates. In the next two sections I will exemplify this pattern by considering initial geminates in Pattani Malay stress and in Trukese Word Minimality. I will then propose that the structure in (5b) is also available word-medially, extending an idea Ham (2001) first proposed. Data from Marshallese will be used to illustrate this pattern.

## 6.3 Initial moraic geminates as moraic onsets ('Initial' geminates word-initially)

### 6.3.1 Pattani Malay and stress

#### 6.3.1.1 The data

Pattani Malay is a dialect of Malay spoken widely among the Muslim communities of the southern provinces of Thailand. The analysis that I will build is based on data

<sup>4</sup> Perhaps though, such a representation could prove useful or necessary in languages where initial geminates are 'weightless' and no re-analysis à la Ham (see below) is available.



presented by Yupho (1989) and will be compared to the analysis Hajek and Goedemans (2003; henceforth H&G) provide for the same set of facts.

The consonantal phonemic inventory of the language includes: the stops /p t c k ʔ b d ʝ g/, fricatives /s h z ɣ/, nasals /m n ɲ ŋ/, post-stopped nasals /m<sup>b</sup> n<sup>d</sup> ɲ<sup>j</sup> ŋ<sup>ʙ</sup>/, liquids /r l/ and the glides /w y/<sup>5</sup>. There are also twelve vowels /i e ε ɪ a u o ɔ ɛ ɶ ʉ ɔ/. Vowel length is predictable, so that vowels in open syllables are long, whereas those in closed syllables are short. Only the schwa-like vowel /ɪ/ is special in that it is relatively short even in open syllables. Consonant length contrasts only word-initially, where the language distinguishes between short consonants and long geminates. Geminates are banned from other positions.

(6) *Initial geminates vs. singletons in Pattani Malay* (Abramson 1999: 592)

<u>Singletons</u>		<u>Geminates</u>	
make	'to eat'	m:ake	'to be eaten'
lamaʔ	'late'	l:amaʔ	'cause to be late'
yato	'comprehensive'	y:ato <sup>6</sup>	'to spread out'
sepaʔ	'to kick'	s:εpaʔ	'to be kicked'
cabe	'branch'	c:abe	'side road'
buŋo	'flower'	b:uŋo	'to bloom'
kukoh	'to be stable'	k:ukoh	'to render stable'

Abramson has attempted to identify the phonetic correlates of geminates in Pattani Malay and has concluded that the primary correlate is the duration of the closure or constriction (Abramson 1987). While generally geminates are three times longer than their singleton counterparts (Abramson 1986), things are a bit more complicated for voiceless stops given that their occlusions are silent, and yet the length distinction is clearly discernible by speakers. In further experimentation, Abramson has discovered that the greater amplitude (Abramson 1991) and higher F<sub>0</sub> (Abramson 1999) that were found in geminates also affect the identification of long vs. short consonants, especially with regard to voiceless stops. However, the picture is not yet complete and further research is required. For our purposes however, it suffices to mention that the distinction between short and long consonants is a real one and is easily identified by speakers.

<sup>5</sup> This is H&G's (2003) description. Yupho does not include the post-stopped nasals among the phonemes, as she considers them clusters of nasals and stops. However, the fact that as she admits, the stop portion is almost inaudible and that they are syllabified in an onset position suggests that we are talking about secondary articulation of a single consonant. Note also that /ɣ/ is a rhotic-like velar (H&G: 83), so it is not clear whether it should be included next to the fricatives.

<sup>6</sup> According to H&G (2003: 83), /ɣ/ cannot geminate. But this example contradicts this statement.

Geminates can be found in loanwords, e.g. *t:ə* from Thai /*tə:n*/ ‘police station’ or much more commonly, they are the product of initial syllable or morpheme reduction. The initial syllable of a word can delete and a free variant is produced where the second onset (in the original form) geminates as in e.g. *buwi* ~ *wi* ‘give’, *sɪdadu* ~ *d:adu* ‘police’, *pɪmatɔ* ~ *m:atɔ* ‘jewellery’ (Yupho 1989: 130). The same effect can also occur in some morphological environments, namely in: i) words that have a derivational or verbal prefix of the type /*Cɪ-*/ plus the stem, ii) reduplicated forms where instead of the reduplicated form, one finds a geminate, iii) cases where a functional word deletes. Some examples with geminates in prefixed forms are given below<sup>7</sup>.

(7) *Geminates in prefixed forms*

<u>Unprefixed form</u>	<u>Prefixed form</u>	<u>Geminate variant</u>
<i>jələ</i> ‘road, path’	<i>bɪjələ</i> ~	<i>j:ələ</i> ‘to walk’
<i>diyi</i> ‘self’	<i>bɪdiyi</i> ~	<i>d:iyi</i> ‘to stand’
<i>kaji</i> ‘no gloss’	<i>mɪŋaji</i> ~	<i>ŋ:aji</i> ‘to study’

Moving on to the stress facts, primary stress is on the final syllable. All preceding syllables receive secondary stress, unless they include the vowel /*ɪ*/ in which case they are stressless. Thus we get:

(8) *Final stress in Pattani Malay*<sup>8</sup>

a.	<i>jələ</i>	‘road path’
	<i>pɛʔdɔh</i>	‘usefulness’
	<i>məkɛŋ</i>	‘food’
b.	<i>sɪpɪnɔ</i>	‘perfect, complete’
	<i>sɪdɔdɪ</i>	‘police’

The data in (a) show that primary stress is placed on the final syllable, and secondary stress on every syllable before that. (b) however reveals the exception to that, namely when the syllable’s vowel is /*ɪ*/, no stress is placed on that syllable.

The interesting bit comes once we consider stress in words with initial geminates, where primary stress retracts from the final syllable and is placed on the

<sup>7</sup> While it seems that geminate formation is not totally oblivious to morphology, data are murky, so not much can be concluded. Moreover, there are still cases, like loanwords and first syllable deletion where no morphology seems to be involved. I will thus assume that gemination is a phonological process which can be morphologically conditioned.

<sup>8</sup> Unlike H&G (2003), but like Yupho (1989), I do not mark vowel length for reasons that should become evident soon. In the mean time, the reader should keep in mind that vowels are long in open syllables, but short in closed ones.

initial one (9a). With the usual caveat about syllables including /ɨ/, the remaining syllables receive secondary stress. In this way, some minimal pairs are also formed (9b).

(9) *Geminates and stress*

	<u>Non-geminated</u>		<u>Geminated</u>
a.	[pimàtɕ]	~	[m:átɕ] ‘jewellery’
	[sɨdàdú]	~	[d:ádù] ‘police’
b.	[bùwɔh] ‘fruit’	~	[b:úwɔh] ‘to bear fruit’ [from /bɨ+buwɔh/]
	[jàlé] ‘road, path’	~	[j:álè] ‘to walk’ [from /bɨ+jalɛ/]

The most startling fact however is the following. According to Yupho, the effect of gemination on stress is so strong that even a syllable with the otherwise weak /ɨ/ will get primary stress if it is preceded by a geminate. Consequently, we have the following contrast.

(10) *Geminates and /ɨ/*

kɨdà	‘shop’	vs.	k:ídà	‘to the shop’	[from /kɨ+kɨdà/]
------	--------	-----	-------	---------------	------------------

In *kɨdà*, the final syllable is stressed as anticipated, while the first syllable carries no stress whatsoever. In *k:ídà* though, it is now the first syllable that includes a geminate which receives primary stress. Secondary stress appears on the final syllable.

In the next section I will briefly summarize Hajek’s and Goedemans’ (2003) attempt to account for these facts by means of moraic onsets. While I share this idea too, I will show that their analysis as a whole presents undesirable implications, which is why I will propose numerous important modifications.

Before I go into this issue, it is worthwhile mentioning that an account that does not make use of gemination seems possible. Examples like the ones in (7) and (9) in particular could perhaps suggest an analysis where the prefix vowel deletes and subsequently a syllabic consonant is formed (Moirá Yip, p.c.). If we were then to argue that the language forms iambs, most of the stress facts would fall out easily.

For instance in a word with an initial singleton, e.g. [diyi], primary stress is final, so the footing with respect to main stress only, i.e. abstracting away from secondary stress for a moment, would be (diyɨ). Its geminated counterpart [di:yi] however would now actually include a syllabic consonant, therefore it would be footed as (ɖ.dí)yi, accounting for the stress pattern described in Pattani Malay above. To also

get the secondary stress facts, it would be crucial to assume that: i) all syllables are bimoraic (i.e.  $CV^{\mu\mu}$ ,  $CV^{\mu}C^{\mu}$  or  $C:V$ ; the latter are in practice disyllables  $C^{\mu}.CV^{\mu}$ ) with the exception of those open ones including /ɪ/ as these would be monomoraic. Recall (see beginning of this section) that this is in line with Yupho's description that all vowels other than /ɪ/ are long in open syllables, and that: ii) the possible iambic feet should either be (H) or (LH) [H=heavy, L=light]. The combination of these allows us to take secondary stress into account too and parse  $[d\dot{\iota}y\acute{f}]$  as  $(d\dot{\iota})(y\acute{f})$  and  $[d:\acute{f}y\grave{\iota}]$  as  $(\acute{f}.d\dot{\iota})(y\grave{\iota})$ . Such analysis however faces some difficulties.

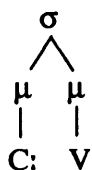
First, from a typological point of view, Bell (1978: 160) argues that in a sample of 53 languages with syllabic consonants, these were found in the following positions in the word: i) medial, ii) initial and medial, iii) medial and final, and iv) all three. Importantly, there were no cases where syllabic consonants emerged in word-initial position only (and for that matter nor in final position only). As we know, Pattani Malay possesses geminates only word-initially. If these geminates were thus considered syllabic consonants, Pattani Malay would be a counterexample to this generalization. Pattani Malay would also be highly unusual in admitting even voiceless stops as syllabic consonants, e.g. as in *c:abε* 'side road', although these are not unprecedented, e.g. Koryak (Bell 1978: 185), Berber (Dell and Elmedlaoui 1985, 1988).

From an empirical point of view, this analysis also fails to explain why only in words with syllabic consonants is primary stress systematically assigned on the first syllable, i.e. the contrast between  $d\dot{\iota}y\acute{f}$  and  $d:\acute{f}y\grave{\iota}$  receives no explanation. In the absence of any supporting evidence for syllabic consonants in Pattani Malay and in the presence of overwhelming phonetic evidence for the existence of initial geminates in Pattani Malay, I opt for an account that builds on the latter.

### 6.3.1.2 Hajek and Goedemans (2003)

H&G (2003) explicitly state that in order that we explain the Pattani Malay data, we need to invoke moraic onsets for the initial geminates since they affect stress assignment. The structure they thus adopt is the following.

#### (11) Initial moraic geminates in Hajek and Goedemans



H&G argue that the conclusion that there can be languages, like Pattani Malay, whose stress placement is determined by moraic onsets is inescapable, if we accept the following claims in phonological theory (most of which are uncontroversial): i) moraic weight affects the stress algorithm in many languages, ii) geminates are generally moraic, and iii) onset geminates exist and can be moraic (cf. Davis 1999). While I agree with this approach, which also happens to be totally compatible with the claims of the present thesis, there is a point in which I differentiate from H&G's conception of moraic onsets. For Goedemans (p.c.), initial moraic geminates call for an analysis that utilizes moraic onsets, but cases like Pirahã do not. The point however is that once we introduce this extra bit of machinery and even if the point of departure in doing so are the initial moraic geminates, then it should be clear that Pirahã and the other similar to it languages could be merely seen as a natural extension to this proposal.

Setting this conceptual issue aside, let us focus on the more technical aspects of H&G's analysis of Pattani Malay. For them, the moraic composition of Pattani Malay syllables is the one below:

(12) *Pattani Malay syllables and their moras in H&G's approach*

- i)  $C\text{ɪ}=1\mu$
- ii)  $CVV=2\mu$ , where V is not /ɪ/
- iii)  $CVC=2\mu$
- iv)  $C\text{ɪ}=2\mu$
- v)  $C:VV=3\mu$ , where V is not /ɪ/<sup>9</sup>

Recall that all vowels apart from /ɪ/ are long in open syllables and short in closed ones. If these are seen as CVV and CVC, then it is reasonable to claim that they are bimoraic, whereas an open syllable with /ɪ/ which does not ever receive stress, could be treated as monomoraic. Geminates on the other hand cause stress attraction, therefore they have to be considered moraic. Thus, when they precede /ɪ/, they render the syllable bimoraic, whereas in front of the other vowels, the syllable becomes trimoraic.

In the examples below, we see that primary stress is word-final when all syllables are maximally bimoraic [(13a.i), (13b.i)]. Any remaining syllables receive secondary stress. In the second examples of each case, a geminate arises which renders

---

<sup>9</sup> As far as I can tell, there is no C:VC example in the data provided, but this should not come as a surprise, since medial clusters are relatively rare. Most of them are homorganic nasals and voiced stops, which actually get to be realized as post-stopped nasals in onset position. Only a few clusters occur in coda-onset position, such as *baʔpɔ* 'why' or *kiyenme* 'sending', but these are usually loanwords or morphologically complex (Yupho 1989: 128-129).

the first syllable trimoraic. By being heavier, it attracts stress on it, which produces the effect of initial primary stress.

- (13) a. i) [bùwóh] (σ)<sub>μμ</sub> (σ)<sub>μμ</sub>  
           ii) [b:úwóh] (σ)<sub>μμμ</sub> (σ)<sub>μμ</sub>  
       b. i) [pimàtós] (σ)<sub>μ</sub> (σ)<sub>μμ</sub> (σ)<sub>μμ</sub>  
           ii) ~ [m:átò] (σ)<sub>μμμ</sub> (σ)<sub>μμ</sub>

Crucially, the analysis up to this point makes the prediction that if the first syllable were bimoraic due to a moraic onset geminate followed by /ɹ/, and the remaining syllables were CVV or CVC, then primary stress would be word-final by virtue of the preference for rightmost stress, since all syllables would be equally bimoraic.

This expectation however is contradicted by the example *k'ídà* (cf. (10)), where primary stress retracts on the initial syllable, although it should stay put on the final. So, how can the initial syllable get stress if it too is bimoraic? The answer H&G give is that Pattani Malay presumably illustrates an instance where onset weight takes priority over general weight. The presence of an onset geminate is enough to attract stress even if this syllable is not trimoraic. To express this fact, they utilize a constraint, whose name is never provided, but is described as follows.

- (14) cf. H&G's (12) on p. 89

The left edge of the main stress foot must be aligned to a moraic consonant<sup>10</sup>

This constraint needs to dominate the constraint that ensures that primary stress will otherwise reside on the final syllable (ALIGN-HEAD-R). No fully-fledged analysis is provided, but it should be clear that H&G's proposal could account for the full range of facts, once the addition of (14) is acknowledged.

The problem with this proposal however relates to the idea of onset weight taking precedence over general weight both on empirical and technical grounds. By claiming that onset weight can take priority over nucleic weight (to simplify), then the proposed pattern is: C:V > CVV. Once this is possible, nothing stops us - in fact it is entirely reasonable - to predict that some other language which is insensitive to onset-weight, but where coda weight matters, would have an analogous pattern, but this time

<sup>10</sup> And this edge must coincide with a moraic onset and not with a moraic coda, because in the latter instance, we would have a violation of syllable integrity.

for codas yielding: CVC > CVV. In other words, coda weight would be more significant than nucleic weight. To my knowledge however, this pattern is unattested.

The patterns we get with respect to CVV and CVC are the following (cf. Gordon 2002 and references cited therein): i) CVV > CVC, CV, i.e. CVV is heavier than both the light CVC and CV (Khalkha Mongolian), ii) CVV, CVC > CV, that is both CVV and CVC are heavy, whereas CV is light (Latin, Yana, Japanese, Finnish) and iii) CVV > CVC > CV where there is a three-way weight hierarchy where CVV is heavier than CVC and CVC is heavier than CV (Kashmiri, Klamath, Chickasaw, Telugu). What we do not get is what H&G predict, namely: CVC > CVV > CV<sup>11</sup>, where the contribution of coda weight is more important than that of the nucleus.

One possible counter-example however is Dutch, which superficially suggests this pattern when one compares words like [xibráltar] and [álma:nàk]. Both end in a closed syllable, which is skipped for (primary) stress purposes. The former has a CVC penult, whereas the latter has a CVV one. And yet, it is only the CVC penult that gets stressed. A CVV one does not. This fact can thus be construed as CVC > CVV, but Gussenhoven (to appear) shows that this is mistaken. Note that the penult vowels here are /a/ and /ɑ/, the difference of which has been claimed to be one of length, i.e. long [a:] vs. short [ɑ]. This point exactly is the one that allows the interpretation of CVC > CVV, since long [a:] is skipped for stress purposes.

However, Gussenhoven argues that the difference instead should be stated as one between tense and lax vowels (van Oostendorp 1995/2000). The idea is then that underlyingly both types of vowels are short and have the same duration in unstressed positions. On the surface, tense vowels will remain short unless they are stressed in which case they lengthen and become bimoraic due to STRESS-TO-WEIGHT; short lax vowels on the other hand can become bimoraic when followed by a coda consonant. Previous works confused the length observed in stressed tense vowels for an underlying property, which is why it was assumed that even when unstressed, they would still be long bimoraic implying that in cases like [álma:nàk], they would need to be considered lighter than the CVC ones. In Gussenhoven's proposal however, this word is actually [álmanàk], with a short unstressed tense vowel in the penult, hence it no longer suggests that CVC syllables act as heavier than CVV ones. These two situations can be

---

<sup>11</sup> Or even CVC > CVV, CV, but then we could argue that vowel length is not really contrastive, so that this actually boils down to CVC > CV which would be possible to get in a language that lacks long vowels. Such a case occurs in Hixkaryana (Hayes 1995), West Tarangan (Gordon 1999) and Moro (Gordon 1999). The latter's weight pattern is CVC > Full V > Reduced V.

summarized below by mentioning the observation that has led each to the corresponding assumptions about the input length.

- (15) *Dutch tense V length*: (I) underlyingly long vs. (II) underlyingly short (Gussenhoven to appear); ' = stressed, ~ = unstressed

	<u>Observation</u>		<u>Input</u>	
(I)	[á:]	therefore	/a:/	suggesting [ǎ:] when unstressed
(II)	[ǎ]	therefore	/a/	suggesting [á:] under stress

As a matter of fact, Gussenhoven explicitly argues that Dutch presents the well-attested CVV, CVC > CV schema, therefore it is no counter-example to the weight generalizations listed above.

The incorrect prediction of CVC > CVV can also associate to another possible pattern. If we have languages where onset weight takes priority, and others where coda weight takes priority, then perhaps we could simultaneously combine the two in one language. This would generate extremely complicated weight systems, such as  $C^\mu V^\mu C^\mu > C^\mu V^\mu$ ,  $V^\mu C^\mu > C^\mu V^\mu$ ,  $V^\mu C^\mu > V^\mu$  which fail to arise.

Finally and on a more technical note, observe that the constraint mentioned in (14), which could perhaps be represented as ALIGN-L (HDFOOT,  $C^\mu$ ) is not only unprecedented, but also quite complex, stuffing in several things at the same time. Thus it does not simply require the alignment of a foot with a consonant, but the alignment of the *head* foot with a consonant which is *moraic*. All this reminds us of a conjunction constraint of some sort. Conjunction has been criticized a lot over the years (see Padgett 2002, Topintzi 2005a for discussion and references cited therein), but still might be needed for certain well-argued cases. In the next section, I will attempt to show that Pattani Malay is not one of these.

By re-examining the data, and drawing slightly different, yet cross-linguistically reasonable assumptions, we can account for Pattani Malay avoiding all the rather unpleasant results of H&G's analysis. We would only need to admit moraic onsets, but this is a move H&G already accept and should by now be well-justified throughout the course of the present work.

### 6.3.1.3 Current analysis and the role of vowels

Recall that a crucial component of H&G's analysis was that CVV and CVC syllables were considered bimoraic, whereas  $C_t$  was monomoraic. This started the problem we



faced when considering a word like [k:ídà]. There, due to the contribution of a mora by the initial onset geminate, both syllables became equally bimoraic, therefore primary stress should be word-final, due to right-edge stress preferences.

But note that this problem only exists if we accept that CVV and CVC are indeed bimoraic. I would like to argue that they are not. As has been mentioned previously on several occasions, vowels in open syllables are long, whereas those in closed ones are short. Effectively this suggests that there is no phonemic vowel length contrast. It could thus be claimed that this lengthening is merely phonetic, but phonologically it has no substance<sup>12</sup>. Syllables of CVV and CVC type would thus phonologically contain just one mora, as it happens in quantity insensitive languages like Standard Modern Greek. As a matter of fact, the claim is that all syllables are monomoraic in Pattani Malay, with the exception of C:V ones, which are bimoraic. I follow H&G in claiming that onset geminates are indeed moraic, therefore they contribute a mora to the syllable. The moraic composition of syllables under this view is the following.

(16) *Pattani Malay syllables and their moras in current approach*

- i) CV(V)=1μ                      for all vowels including /ɨ/
- ii) CVC=1μ
- iii) C:V=2μ                      for all vowels including /ɨ/

The advantage of this proposal is that it can straightforwardly account for all the examples, including *k:ídà*. To see how, let us re-examine *k:ídà* and the other relevant data in the new light.

- (17) a. *Non-geminate, non-/ɨ/ words*  
           [bùwɔ́h]            (ð)μ (ɔ́)μ
- b. *Non-geminate, /ɨ/ words*  
           [pimàtɔ́]            σμ (ð)μ (ɔ́)μ
- c. *Geminate, non-/ɨ/ words*  
           [b:úwɔ́h]            (ɔ́)μμ (ð)μ
- d. *Geminate, /ɨ/ words*  
           k:ídà                (ɔ́)μμ (ð)μ

<sup>12</sup> The central vowel /ɨ/ does not lengthen at all, as is also the case in Yakima Sahaptin (Hargus 2001 and below). This may relate to its inherent weakness (van Oostendorp 1995/2000) or its low sonority (de Lacy to appear; see footnote below for more discussion). /ɨ/ is commonly schwa-like, as it seems to be the case in Pattani Malay too. The absence of long /ɨ/ is then fully parallel with the lack of long /ə/ that other languages exhibit, e.g. Yiddish (Albright 2002: 7) and Yupik (Bakovic 1996: 13).

In (c)-(d), a geminate is present and correctly attracts stress, because the syllable that includes it is bimoraic. All remaining syllables are monomoraic. Consequently, the use of the widely-accepted WSP, without any additional constraints such as the one that H&G utilize, i.e. “The left edge of the main stress foot must be aligned to a moraic consonant” (cf. (14)) will generate the right result provided it will be ranked above ALIGN-HEAD-R, which requires default final stress.

The new problem now is that if all syllables other than C:V are monomoraic, then how come syllables with a reduced vowel /ɪ/ never receive (secondary) stress, whereas those with full vowels do (compare (17a) with (17b))? I will argue that this property does not relate to the moraic content of the syllable, but with the fact that central vowels are cross-linguistically weak and low in terms of sonority. It is well-known that the stress systems of several languages recognise this property and as a result avoid assigning stress on such vowels. Pattani Malay will merely need to be added to the pool of such languages.

Kenstowicz (1994b) presents Mordwin, Kobon, Chukchee, Aljutor and Mari as cases where sensitivity to the relative sonority of the vowels can be detected. Although the pattern of each of these is somehow different from the Pattani Malay one (and will not be reviewed here), the common property is that all these languages follow the hierarchy below:

(18) Kenstowicz’s (1994b) Peak Hierarchy:

\*P/ə >> \*P/i, u >> \*P/e, o >> \*P/a

Reference to \*P/ə is more general and encompasses other central schwa-like vowels such as /ɪ/ as mentioned with regard to Kobon and Mari. This hierarchy explicitly states that there will be languages which will avoid stressing a schwa-like vowel. The same point is made in Gordon (2002) who uses Chuvash and Javanese as representative examples. Gordon’s (2002: 71) remark that: “Virtually all weight distinctions based on vowel quality, including distinctions between short central vowels and other vowels, occur in languages in which vowel length is not phonemic” is also interesting and compatible with what we find in Pattani Malay. This observation makes the possibility that there is no real weight contrast coming from vowels all the more likely.

An analogous situation arises in Yakima Sahaptin, as reported in Hargus (2001). Sahaptin has four vowels /i ɪ u a/, all of which have long counterparts with the exception of /ɪ/, which cannot appear long. There is also a strong trend for /ɪ/ to remain unstressed,

just like in Pattani Malay. Hargus, following van Oostendorp (1995/2000) also observes that Sahaptin /ɨ/ has other properties which reveal its schwa-like nature, such as the fact that /ɨ/, in contrast to the other vowels, deletes, fails to appear in diphthongs and does not emerge word-finally. As far as I have been able to tell given Yupho's (1989) description of Pattani Malay, this language too presents most of these properties. There are no examples with a word-final /ɨ/ and /ɨ/ tends to delete in a word-initial position (cf. (9)), although occasionally other vowels can delete in the same position, e.g. [b:aeʔ] from /baeʔ-baeʔ/ (Yupho 1989: 131). According to Yupho, there are no diphthongs in the language, so that property cannot be tested. All in all then, the marked nature of /ɨ/ and its inability to get stressed find precedents in several other languages.

Finally, de Lacy (to appear) revisits this phenomenon from a broader perspective and shows how we can formally capture the fact that languages whose stress algorithm is sonority-driven, group vowels in certain ways based on their sonority profile (conflation). Conflation can group any set of contiguous categories so that they behave in the same way. Thus given the sonority hierarchy:  $a > \varepsilon, \text{ɔ} > e, o > i, u > \text{ə} > \text{ɪ}$ , it is possible for a language to favour stressing the contiguous peripheral low and mid vowels over the contiguous central and high peripheral ones, i.e.  $a, e, o > i, u, \text{ə}$  (Nganasan). It is not possible however to have a system where it is better to stress mid vowels over low and high ones, i.e.  $e, o, \text{ə} > a, i, u, \text{ɪ}$ , since peripheral mid vowels are not contiguous with the mid central ones (and similarly for the low and high ones).

While the reader should consult de Lacy for details on how conflation works, it is sufficient to mention that this mechanism can easily generate a language like Yil where the stress preference is:  $a, i, u, e, o > \text{ə}$ . In other words, Yil avoids stressing the mid central vowel /ə/. This situation is exactly analogous to Pattani Malay, which avoids stressing the high central vowel /ɨ/. It should thus be clear that this preference is cross-linguistically well-grounded<sup>13</sup>. In the interests of simplicity and since this detail is tangential to the broader picture of Pattani Malay I would like to draw at this point, I will use Kenstowicz's constraint \*P/ɨ to militate against a stressed /ɨ/ instead of de Lacy's more elaborate equivalent.

---

<sup>13</sup> While for de Lacy it is the low sonority of central vowels that sometimes makes them unable to bear stress, for other researchers, such as van Oostendorp (1995/2000), it is their weakness due to the lack of featural content. In Pattani Malay, either approach would work, since it is only /ɨ/ that behaves differently from other vowels. However, as de Lacy observes, if featurelessness is responsible for this effect, then in a language like Nganasan where all of /ɨ ə i u y/ equally repel stress, we would need to say that all these vowels lack features. But this would make them phonologically indistinguishable, an undesirable result.

We are now ready to proceed to the analysis which will make use of the above constraints ALIGN-HD-R and \*P/ɿ combined with some additional ones to get the full range of Pattani Malay facts.

Let us first take the simplest case of a Pattani Malay word, namely one which contains any vowels other than /ɿ/ and which lacks an initial geminate, e.g. /buwɔh/ or /mækɛŋ/. As we have previously seen, such words receive final primary stress and secondary stress on the remaining syllables, i.e. [bùwɔ́h] and [màkɛ̀ŋ]. In other words, every syllable receives some level of stress. This property is rather strange and unusual, since languages usually avoid consecutive stresses which entail numerous \*CLASH violations.

On the other hand, if Yupho's (1989) description is accurate, then there seems to be a contrast, which needs to be accounted for, between syllables with full vowels that receive secondary stress and those with an /ɿ/ that lack stress altogether, e.g. [mà] and [pi] respectively in [pimàtɔ́]. Still then, it is an issue to what extent this is really phonological stress, rather than perhaps a phonetic effect due to the relatively longer duration of a CVC or a CVV lengthened syllable compared to shorter Cɿ. In the absence of any conclusive evidence, I will build an analysis that assumes and accounts for this contrast.

The first step in doing so is to explain how all syllables receive stress and thus form monosyllabic feet. The answer to this lies in the following foot alignment constraints.

(19) ALIGN-L (Ft, FtHd): Align the L edge of every foot with the L edge of a foot head

ALIGN-R (Ft, FtHd): Align the R edge of every foot with the R edge of a foot head<sup>14</sup>

As Prince (1997, citing Bruce Tesar) and Green (2002) observe, the constraints above can both be simultaneously satisfied only in a monosyllabic foot. Prince (1997) also notes that if these constraints are very highly-ranked, then we can produce languages which only allow monosyllabic feet, like Cantonese (Maira Yip, p.c). Obviously, feet in Pattani Malay, which are strictly monosyllabic as the stress facts reveal, call for a similar account. In addition, all syllables - with the exception of those including /ɿ/ (we


---

<sup>14</sup> Following Prince (1997), I will abbreviate these as Ft-Hd-L and Ft-Hd-R respectively.

will return to this in a moment) - are parsed into feet. Thus, Pattani Malay feet require the ranking in (20) which expresses the fact that all syllables have to be parsed in monosyllabic feet, even at the expense of creating several stress clashes or non-binary feet. Tableau (21) exemplifies this result.


(20) *Monosyllabic feet in P. Malay*: Ft-Hd-L, Ft-Hd-R, PARSE- $\sigma$  >> FTBIN, \*CLASH

(21) [bùwóh]: Ft-Hd-L, Ft-Hd-R, PARSE- $\sigma$  >> FTBIN, \*CLASH

	buwóh	Ft-Hd-L	Ft-Hd-R	PARSE- $\sigma$	FTBIN	*CLASH
a.	(buwóh)	*!				
b.	(búwóh)		*!			
c.	bu(wóh)			*!		
 d.	(bù)(wóh)				**	*

To keep tableaux as manageable as possible, I will use the cover constraint FTFORM to represent the constraints Ft-Hd-L, Ft-Hd-R, PARSE- $\sigma$ . While (21d) is the correct winner, it is not the only possible candidate with monosyllabic feet. [(bú)(wóh)] is another. To rule this out, we need to make reference to ALIGN-R (PrWd, HdFt) [abbreviated as ALIGN-HD-R], which requires that primary stress appears at the right edge. Incorporating this to our ranking - although no ranking argument can be established - provides the right candidate.

(22) [bùwóh]: FTFORM, ALIGN-HD-R

	buwóh	FTFORM	ALIGN-HD-R
a.	(bú)(wóh)		*!
 b.	(bù)(wóh)		

The next ingredient in the analysis involves the treatment of words including /ɨ/. The constraint \*P/ɨ now comes handy. If we rank \*P/ɨ >> FTFORM, then we ensure that the syllable which includes /ɨ/ will be left unstressed. An illustration with the example [pimàtɔ] follows. Here, we get primary stress on the final syllable and secondary on the syllable before. The first syllable remains unstressed. This can be done by either leaving it unparsed (23e) or by parsing it in the foot tail position (23d). Any of the two will do, as empirically they are the same.

(23) [pimàtɔ́]: \*P/ɪ >> FtFORM, ALIGN-HD-R

	pimato	*P/ɪ	FtFORM	ALIGN-HD-R
a.	(pǐ)(mà)(tɔ́)	*!		
b.	pǐ(matɔ́)		* (Ft-Hd-L), *! (PARSE-σ)	
c.	pǐ(mà)(tɔ́)		* (PARSE-σ)	*!
d.	(pimà)(tɔ́)		* (Ft-Hd-L)	
e.	pǐ(mà)(tɔ́)		* (PARSE-σ)	

Summing up the preceding discussion, we have seen that the part of Pattani Malay stress which does not involve any weight considerations - since all syllables are monomoraic - requires the ranking:

(24) *Pattani Malay stress (to be revised)*

\*P/ɪ >> Ft-Hd-L, Ft-Hd-R, PARSE-σ, ALIGN-HD-R >> FtBIN, \*CLASH

What this means is that primary stress is final (ALIGN-HD-R) and all remaining syllables receive secondary stresses (PARSE-σ) forming monosyllabic feet (Ft-Hd-L/R >> FtBIN). The only exception to that is with a syllable that comprises the schwa-like /ɪ/, which has to stay unstressed (top-ranked \*P/ɪ).

Now we only need to consider cases which involve an initial geminate. Let us begin with an example like [b:úwə̀h], which is very similar to [bùwóh] previously examined in (21) and (22). The former differs in that it includes a geminate, whose effect is to attract *primary* stress on its host syllable. This can be attributed to a high-ranking WSP constraint whose effects now become visible, since it applies on a syllable which is bimoraic by virtue of one mora of the vowel and the one of the initial moraic onset geminate.

However, note that it is primary stress which shows up on this syllable, therefore reference to general WSP is not sufficient, as it can be satisfied by either primary or secondary stress. What we need is a version of this constraint specific to primary stress, namely WSP<sub>PrWd</sub> (cf. McGarrity 2003).

(25) WSP<sub>PrWd</sub>: Heavy syllables receive primary stress<sup>15</sup>

<sup>15</sup> WSP can still be present in the ranking, but no matter its ranking with respect to WSP<sub>PrWd</sub>, the winning form will be the one that assigns primary stress on the initial heavy syllable, so it will always satisfy both WSP<sub>PrWd</sub> and WSP, whereas secondary stress on that syllable will satisfy WSP, but violate WSP<sub>PrWd</sub>. For this reason, I will leave WSP out of the following tableaux.

With this in mind, the only candidates we need to take into account are: i) the one that assigns primary stress on the first syllable (26a), and ii) the one that assigns secondary stress on the same syllable (26b). Due to the ranking  $WSP_{PrWd} \gg ALIGN-HD-R$ , the winner will be the former candidate, despite misaligning primary stress from the right edge.

(26)  $WSP_{PrWd} \gg ALIGN-HD-R$

	buwɔh	$WSP_{PrWd}$	$ALIGN-HD-R$
☞ a.	(b:ú)(wɔ́h)		*
b.	(b:ù)(wɔ́h)	*!	

Only one thing remains to be discussed<sup>16</sup>. What happens when there is a word that consists of both a geminate and a /ɪ/? This is exactly the [k:ɪdà] case which posed problems for H&G and required the introduction of undesirable constraints such as (14). Given that the first syllable can be stressed despite the presence of /ɪ/ indicates that the effect of geminates and the activation of  $WSP_{PrWd}$  is overwhelming, therefore it must also dominate  $*P/\text{ɪ}$ .

(27)  $WSP_{PrWd} \gg *P/\text{ɪ}$

	k:ɪda	$WSP_{PrWd}$	$*P/\text{ɪ}$
☞ a.	(k:ɪ)(dà)		*
b.	k:ɪ(dá)	*!	

With this modification, the newly acquired ranking for Pattani Malay stress is given below.

(28) *Pattani Malay stress (final version)*

$WSP_{PrWd} \gg *P/\text{ɪ} \gg Ft-Hd-L/R, \text{PARSE-}\sigma, \text{ALIGN-HD-R} \gg FtBIN, *CLASH$

<sup>16</sup> There is actually a Richness of the Base issue here involved too. What happens with underlying geminate consonants in word-medial/final position? We know that these do not emerge, therefore some markedness constraint militating against them must be highly-ranked. In the interests of simplicity - as this would shift our focus here which is the initial geminates - but not of elegance, I assume that something like the ad hoc constraint  $*MEDIAL/FINAL \text{ GEMS}$  is in operation.

#### 6.3.1.4 Pattani Malay summary

In this part of the chapter, I have examined the Pattani Malay stress pattern as described in Yupho (1989) and proposed an alternative analysis to the one argued for in Hajek and Goedemans (2003). Both accounts however share the same important intuition, namely that the initial moraic geminates of the language should be represented as moraic onsets, thus providing an additional case of onset weight.

Nonetheless, there has been disagreement over the specifics of the analysis. Thus, while H&G explicitly state that onset weight can take priority over general weight, I have argued that this is undesirable for reasons both empirical and technical. Instead I have pursued the idea that the language only has a weight contrast with regard to consonants, so that otherwise syllables are monomoraic. Whenever all syllables are monomoraic, primary stress is word-final due to alignment considerations. Whenever there is a bimoraic syllable - which is necessarily initial, given that geminates occur word-initially only - this attracts stress, because of high-ranking WSP. Finally, to account for the absence of stress on syllables with the vowel /ɨ/, I have made reference not to weight but rather to quality considerations, since it is cross-linguistically observed that central vowels commonly remain unstressed. Conforming to the WSP however is more important than leaving a syllable with /ɨ/ unstressed, which is why we can find stressed /ɨ/'s only when preceded by a geminate as in *k:ɨdà*.

#### 6.3.2 Trukese Word Minimality

The next language I will consider is Trukese, a Micronesian language, spoken on the islands of the Truk Atoll south of Guam. Trukese is similar to Pattani Malay in that it too presents initial moraic geminates, which I will argue can likewise be represented as moraic onsets. This time however, the effects of geminates will be identified with respect to Word Minimality and not stress.

Minimal Words consist of either a long vowel, i.e. (C)VV or a geminate plus a short vowel, i.e. C<sub>i</sub>C<sub>i</sub>V. CVC or CV words are not allowed (Davis 1999)<sup>17</sup>, which suggests that coda consonants are not moraic (Muller 1999). In fact, Davis' and Torretta's (1998: 112, fn. 2) comment that the only codas available are the first part of geminates, except in word-final position and Muller's (1999: 394) statement that no

---

<sup>17</sup> Although Muller (1999: 395, fn. 3, and p. 394, example in (2)) observes that there are verbs which are monomoraic, such as *ma* 'to be ashamed' or *kak* 'ring'. This is never the case for nouns.



consonant clusters occur in Trukese lead us to the conclusion that singleton codas only occur word-finally and geminate codas only occur word-medially. Given the minimality and the weight facts below, I will also argue that singleton codas are not moraic, but geminate ones are. But let us first exemplify the minimality pattern with some examples.

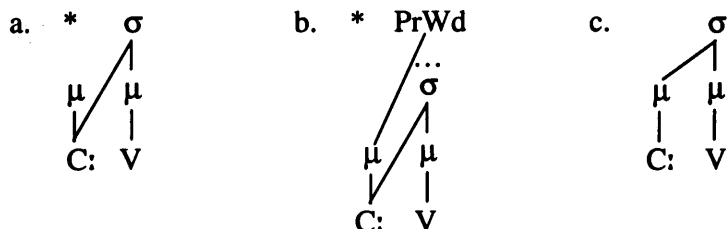
(29) *Trukese Minimal Words*

	<u>Form</u>	<u>Gloss</u>
a. CVV words	maa	'behaviour'
	oo	'omen'
	tée <sup>18</sup>	'islet'
	núú	'unripe coconut'
b. CCV words	tto	'clam sp.'
	kka	'taro sp.'
	čča	'blood'

This pattern suggests that a Word Minimality restriction is applicable in Trukese. A bimoraic minimum is imposed, which can either be satisfied by a long vowel or by a geminate followed by a short vowel. Geminates consequently contribute to weight. Additional data (see below) highlighting an interesting interaction between final deletion and lengthening also illustrate that geminates are indeed moraic.

The next question has to do with the correct representation for such a consonant. If we are right that representations like (4) and (5a) mentioned previously are not appropriate, then we are left with (5b), i.e. the representation of an initial geminate as a moraic onset [structures repeated as (30a), (30b) and (30c) respectively].

(30) Possible moraic representations of initial geminates



In what follows, I will briefly summarize the analyses Davis and Torretta (1998) and Muller (1999) suggest and show how they are compatible with the current proposal. I will start by describing the data where the interaction between vowel deletion and minimality considerations relate to lengthening.

<sup>18</sup> é=[ʌ], ú=[ɪ] (Davis and Torretta 1998: 112, fn. 2).

In Trukese, there is a pervasive process under which the root-final mora deletes if it appears word-finally. As a result of this, one can see effects that interact with Word Minimality. In particular, when a trimoraic word shows deletion of its final mora, nothing extra happens, because the resulting word still satisfies the bimoraic word minimum (31).

(31) *Final mora deletion in words of more than or equal to 3μ*

<u>Suffixed form</u>		<u>Unsuffix form</u>	
sawaa-n	'taro of'	sawa	'taro'
orosseti-n	'shore of'	orosset	'shore'

The first column shows suffixed forms, where deletion fails to apply, because the final mora is not in word-final position. It is protected by the suffix /-n/ which is non-moraic. On the contrary, unsuffixed forms present mora deletion. This is why the long vowel becomes short in /sawaa/ and the short vowel deletes altogether in /orosseti/. The resulting form is minimally bimoraic, therefore no other process occurs. Things are different in smaller words. In particular, if the original word is bimoraic, deletion leads to a monomoraic word. Since this is not permitted, lengthening occurs, which can take the guise of either lengthening of a vowel (32) or consonant gemination (33).

(32) *Final mora deletion causing V-lengthening*

<u>Suffixed form</u>		<u>Unsuffix form</u>	
a. fasa-n	'nest of'	faas	'nest' *fas
b. fæne-n	'building of'	fææn	'building' *fæn

(33) *Final mora deletion causing C-gemination*

<u>Suffixed form</u>		<u>Unsuffix form</u>	
a. fitta-m <sup>w</sup>	'your package'	ffit	'package' *fiit, *fitt
b. kikki-k	'you move'	kkik	'move' *kiik, *kikk

In (32), we see that the anticipated unsuffixed form *fas* does not occur, because it is monomoraic. Vowel lengthening takes place yielding the bimoraic *faas*. The same result is achieved in (33), but this time in a different way; consonant gemination occurs, but rather than being realized word-finally as in *fitt*, it happens word-initially, thus we get *ffit*. The reason for this is that word-final geminates are never admitted in Trukese. Such 'long-distance' gemination has been dubbed "geminate throwback" (33) and obviously corroborates the finding about the geminates' moraicity since no additional V-lengthening is required to fulfill the bimoraicity requirement.

The next question is of course how to account for these facts, given that there is no obvious reason why some roots choose V-lengthening and some others C-gemination. Davis and Torretta (1998) impose a markedness constraint  $*\mu_{\#}$  (or alternatively named FREE- $\mu$  = ‘the final mora of the input cannot be realized in word-final position in the output’) that forces deletion of the final mora<sup>19</sup>, and lengthening is imposed only whenever applicable, that is to avoid minimality violations<sup>20</sup>. For instance, a form like /o<sup>μ</sup>mo<sup>μ</sup>su<sup>μ</sup>/ becomes [o<sup>μ</sup>mo<sup>μ</sup>s] without any lengthening establishing  $*\mu_{\#} \gg \text{MAX-}\mu$ . The alternative [o<sup>μ</sup>mo<sup>μμ</sup>s] is not chosen, because lengthening of underlying short /o/ would needlessly violate IDENT- $\mu$ , which requires that corresponding segments have identical weight<sup>21</sup>. Thus, IDENT- $\mu \gg \text{MAX-}\mu$ <sup>22</sup>. Similarly, a word with an initial geminate such as /f<sup>μ</sup>é<sup>μ</sup>ne<sup>μ</sup>/ ‘advice’ surfaces as [f<sup>μ</sup>é<sup>μ</sup>n] without any other modifications. However, in a form like /ti<sup>μ</sup>pe<sup>μ</sup>/, the output [ti<sup>μ</sup>p] would be subminimal, which is why lengthening applies yielding [ti<sup>μμ</sup>p]. IDENT- $\mu$  is violated, since short /i/ lengthens, but WDMIN is a higher priority, i.e. WDMIN  $\gg$  IDENT- $\mu$ . WDMIN is also a higher priority than  $*\mu_{\#}$  because a word like /ma<sup>μμ</sup>/ ‘behaviour’ will stay [ma<sup>μμ</sup>] rather than shorten to the subminimal \*[ma<sup>μ</sup>]. All these lead to the ranking: WDMIN  $\gg$   $*\mu_{\#} \gg$  IDENT- $\mu \gg$  MAX- $\mu$ .

The interesting point arises in surface suffixed forms like [múnnú-n] from input /mú<sup>μ</sup>n<sup>μ</sup>ú<sup>μ</sup>-n/ ‘upper back [-n=relational]’, whose bare root could in principle emerge either as *mmún* [m<sup>μ</sup>ú<sup>μ</sup>n] with initial consonant gemination, *múnn* [mú<sup>μ</sup>n<sup>μ</sup>] with final consonant gemination or as *múún* [mú<sup>μμ</sup>n] with vowel lengthening<sup>23</sup>. Final consonant gemination is easily excluded because of a top-ranked constraint in Trukese against final geminates (\*FINAL GEM). Each of the remaining two forms violates IDENT- $\mu$  twice: *mmún* because it lengthens *m* and shortens *n* and *múún* because it shortens *n* and lengthens the first *u*. In each case, the final mora is lost, so MAX- $\mu$  is equally violated.

<sup>19</sup> Some constraints are shared among Davis and Torretta (1998) and Muller (1999), but they choose to use different names for them. Given that I consider Muller’s analysis slightly more advantageous, I will use Muller’s constraints for consistency. The equivalent constraints are the following: Free- $\mu$  and Weight-Identity in Davis and Torretta correspond to Muller’s  $*\mu_{\#}$  and IDENT- $\mu$  respectively.

<sup>20</sup> In the following discussion I assume the following standard moraic structure: short V=1 $\mu$ , long V=2 $\mu$ , short C=0 $\mu$ , geminate C=1 $\mu$ .

<sup>21</sup> Davis and Torretta make no reference to any DEP- $\mu$  violations, but presumably they assume it is not violated here or in other similar cases below, where the total number of moras between input and output remains the same.

<sup>22</sup> A further ranking argument  $*\mu_{\#} \gg$  IDENT- $\mu$  can also be established, but explaining how would take us too far afield (see Davis and Torretta for details).

<sup>23</sup> [mu<sup>μ</sup>n<sup>μ</sup>] with a final singleton moraic consonant is not considered, but presumably it is banned, because singleton moraic codas are not permitted.

As Davis and Torretta (1998) show that these are the only two surviving candidates, they need to be able to somehow select the actual winner *mmún* with “geminate throwback”. To do so, they stipulate the ranking  $\text{MAX-}\mu_c \gg \text{MAX-}\mu_v$  which gives priority to retaining input consonantal moras on consonants on the surface over the corresponding requirement for vocalic moras. The net effect is that  $[\text{m}^{\text{H}}\text{u}^{\text{H}}\text{n}]$  fares better than  $[\text{m}\text{u}^{\text{H}}\text{n}]$  because it manages to retain one vocalic and one consonantal mora of the input  $/\text{m}\text{u}^{\text{H}}\text{n}\text{u}^{\text{H}}/$ , whereas  $[\text{m}\text{u}^{\text{H}}\text{n}]$  retains both vocalic moras, but has no consonantal correspondent for the consonantal mora  $/\text{n}^{\text{H}}/$ . The ranking  $\text{MAX-}\mu_c \gg \text{MAX-}\mu_v$  ensures that this proves fatal.

While the ranking imposed seems to yield the right effects, there are certain shortcomings in this approach. First, it makes crucial use of the distinction between vocalic and consonantal moras, which is not unreasonable, but is not well-supported either. Also the ranking  $\text{MAX-}\mu_c \gg \text{MAX-}\mu_v$  seems rather controversial, as one would perhaps expect higher faithfulness rank for vocalic moras rather than consonantal ones. The biggest problem however that this analysis suffers from is that it predicts that suffixed forms surfacing with a medial geminate as in *múnnú-n* should include a medial geminate in the input, i.e.  $/\text{CVC}_i\text{C}_i\text{V}/$  and produce forms with ‘gemination throwback’ like *mmún* when unsuffixed. Muller (1999) though points out that there are other suffixed words with medial geminates, which when unsuffixed present vowel lengthening instead of consonant gemination. The relevant contrast is the following:

(34) *Roots with surface medial geminates when suffixed*

<u>Suffixed form</u>		<u>Unsuffixed form</u>		
a. <i>fitta-m<sup>w</sup></i>	‘your package’	<i>ffit</i>	‘package’	*fiit, *fitt
b. <i>tappa-m<sup>w</sup></i>	‘your coconut’	<i>taap</i>	‘coconut’	*ttap, *tapp

While Davis and Torretta (1998) do not consider this contrast, an extension of their analysis on roots such as  $[\text{tappa-m}^w]$  would be bound to incorrectly produce unsuffixed *\*ttap* instead of *taap*. Muller tackles this problem by drawing attention to the following morphological alternations.

(35) *Morphological alternations and geminates*<sup>24</sup>

<u>Prefixation</u>	<u>Suffixation</u>	<u>No affixation</u>
<i>initial geminate</i>	<i>medial geminate</i>	<i>initial geminate</i>
a. <i>o-kk<sup>w</sup>un</i> ‘rotate it’	<i>kunnu-n</i> ‘its rotation’	<b>kkun</b> ‘rotate’

<sup>24</sup> To highlight the length alternations involved in (35), I mark corresponding consonants in boldface.

$\text{æ-ppɛp}$ ‘cause to skip’ <i>initial singleton</i> b. <i>fita-tap</i>	<i>peppa-n</i> ‘its skipping’ <i>medial geminate</i> <i>tappa-n</i>	<i>ppɛp</i> ‘skip’ <i>long vowel (nouns)</i> <i>taap</i>
---	---	--

In (35a) we find alternations between e.g. [k]~[kk] and [n]~[nn]. In contrast, in (35b), only *p* alternates between [p]~[pp], but [t] merely appears as a singleton. Note however that in the suffixed form - the one Davis and Torretta use to construct the input - all words have an identical [CVC<sub>i</sub>C<sub>i</sub>V] pattern. The alternations above cannot be accounted for unless there is some kind of intermorphemic gemination process, for which no evidence is available. Moreover, under the assumption that inputs are consistently /CVC<sub>i</sub>C<sub>i</sub>V/, as Davis and Torretta suggest, we cannot explain that in the first word, sometimes /k/ geminates and sometimes /n/. Muller’s response therefore is that there is no single input. Her generalisation is that nouns and verbs with geminate alternations (35a) have an input with two geminates, i.e. /C<sub>i</sub>C<sub>i</sub>VC<sub>j</sub>C<sub>j</sub>V/, whereas nouns that show vowel lengthening (35b) have an input with a single geminate /CVC<sub>i</sub>C<sub>i</sub>V/<sup>25</sup>.

In her formal analysis, Muller too uses high-ranking constraints such as WDMIN, \*FINAL GEM and \*μ<sub>#</sub><sup>26</sup>. All these outrank IDENT-μ. Furthermore, Muller assumes, without further justification, that initial geminates are indeed moraic onsets. For this reason, she also uses low-ranked \*MORAIC ONSET which effectively allows moraic onsets given the opportunity. The tableaux below show how the analysis works. Throughout, I will adopt the following convention: to facilitate candidate evaluation, I will present the moraic structure of input and candidates, followed by the forms in italics to clarify their length, i.e. represent long segments as doubled and short ones as single. The offending segment violating IDENT-μ is presented in brackets whether it implies lengthening of an underlyingly short segment or the reverse.

(36) *Vowel lengthening in inputs with a single geminate*

	/ta <sup>μ</sup> p <sup>μ</sup> a <sup>μ</sup> /	<i>tappa</i>	*FINAL GEM	WDMIN	*μ <sub>#</sub>	IDENT-μ	*MORAIC ONSET
a.	ta <sup>μμ</sup> p	<i>taap</i>				*(a) *(p)	
b.	t <sup>μ</sup> a <sup>μ</sup> p	<i>ttap</i>				*(t) *(p)	*!
c.	ta <sup>μ</sup> p	<i>tap</i>		*!		*(p)	
d.	ta <sup>μ</sup> p <sup>μ</sup>	<i>tapp</i>	*!				
e.	ta <sup>μ</sup> p <sup>μ</sup> a <sup>μ</sup>	<i>tappa</i>			*!		

<sup>25</sup> Of course, there are inputs without any geminates at all, such as (32).

<sup>26</sup> I have adapted Muller’s original \*COMPLEX constraint to \*FINAL GEM so that its meaning and role is clear.

In both tableaux (36) and (37), candidates that present final geminates [(36d), (37b)], lack final mora deletion [(36e), (37c)] or fail Word Minimality requirements [(36c)] are ruled out by high-ranking constraints. The burden then falls on IDENT- $\mu$  and \*MORAIC ONSET. The remaining candidates in (36), i.e. (a) and (b) fare equally with respect to IDENT- $\mu$ . Both shorten an underlyingly geminate consonant, but in addition to this, the former lengthens a short vowel, whereas the latter lengthens a short consonant. Since IDENT- $\mu$  does not select a winner, the decision is passed onto \*MORAIC ONSET which chooses (36a) with V-lengthening, because the contender (36b) violates this constraint by having an initial geminate.

In (37), the situation changes. This time the input includes two geminates, thus IDENT- $\mu$  violations are different. The candidate with V-lengthening (37d), incurs extra violations of IDENT- $\mu$  when compared to (37a) because not only does it shorten the first geminate, but it also lengthens the vowel. In contrast, the winner merely shortens the second geminate, a process that (d) is also bound to present given that final geminates are impermissible. The output thus correctly includes an initial geminate and lacks any extra process.

(37) *Vowel lengthening in inputs with two geminates*

	<i>/p<sup>μ</sup>ε<sup>μ</sup>p<sup>μ</sup>a<sup>μ</sup>/ ppeppa</i>	*FINAL GEM	WD MIN	*μ <sub>#</sub>	IDENT-μ	*MORAIC ONSET
a.	p <sup>μ</sup> ε <sup>μ</sup> p ppep				*(p)	*
b.	p <sup>μ</sup> ε <sup>μ</sup> p <sup>μ</sup> ppepp	*!				*
c.	p <sup>μ</sup> ε <sup>μ</sup> p <sup>μ</sup> a <sup>μ</sup> ppeppa			*!		*
d.	pε <sup>μμ</sup> p peep				*(p)* <sub>(ε)</sub> *(p)!	

An additional detail has to be considered to cover a fuller range of facts. It is expected that inputs like /ppeppa/ should surface with two geminates when suffixed, because the second geminate would no longer be word-final. This prediction however is falsified by the data, as the example /ppeppa-n/ → [peppan] ‘skipping of’ illustrates. More extended observation leads to the conclusion that no word is allowed to include two geminates in the word, a fact that has been interpreted by both Davis and Torretta (1998) and Muller (1999) as an OCP effect, banning the co-occurrence of multiple geminates within a single morpheme (OCP GEM<sub>{MORPHEME}</sub>)<sup>27</sup>. The result is that one of the geminates de-

<sup>27</sup> This is not an inter-morphemic restriction, cf. *kučču-ηηaw* ‘to fit badly’ (Muller 1999: 403) which is well-formed. The OCP explanation is not particularly appealing since the OCP usually makes reference to tones or features, but not to more abstract notions such as weight. In fact, it is difficult to imagine how

geminate, but which? It can be shown that it is the first one that consistently does so. Muller claims that this effect can be achieved by the constraint \*MORAIC ONSET, which whenever possible, will exclude candidates with moraic onset geminates. While I argue that this is the correct solution, one needs to be a bit more specific in the syllabification assumed.

Given that Muller allows geminate syllabification wholly in the onset, it is conceivable to assume the same for a medial geminate. But this would not be permitted due to the action of \*MORAIC ONSET, which of course would penalise candidates with onset geminates. There is however one way to keep one of the geminates and avoid violating \*MORAIC ONSET, but this also requires the consideration of NOCODA which has to be dominated by \*MORAIC ONSET.

(38) *Suffixed input forms with two geminates*

	$/p^{\mu}\epsilon^{\mu}p^{\mu}a^{\mu}+n/$	$ppeppa+n$	OCP GEM <sub>{MORPH}</sub>	IDENT- $\mu$	*MORAIC ONSET	NO CODA
a.	$p\epsilon^{\mu}p^{\mu}a^{\mu}n$	$pep.pan$		*		*
b.	$p\epsilon^{\mu}p^{\mu}a^{\mu}n$	$pe.ppan$		*	*!	
c.	$p^{\mu}\epsilon^{\mu}pa^{\mu}n$	$ppe.pan$		*	*!	
d.	$p^{\mu}\epsilon^{\mu}p^{\mu}a^{\mu}n$	$pppe.pan$	*!			*

The last candidate loses as it violates the OCP. The remaining candidates equally violate IDENT- $\mu$ , thus the decision falls onto \*MORAIC ONSET. Candidates (b) and (c) have just one geminate, but both syllabify it in the onset, thus being penalised by \*MORAIC ONSET. There is yet another candidate which avoids such violations as shown in (a). This time, the medial geminate takes the usual flopped structure of geminates and only violates low-ranked NOCODA. What this grammar of Trukese suggests then is that while initial geminates are analysed as moraic onset geminates, medial ones are of the traditional kind, where the consonant spans the coda and onset.

Taking this statement a step further, we can argue that only the codas of geminates are moraic due to their underlying moraity. On the other hand, singleton codas are not as the minimality facts suggest, presumably because the constraint which assigns coda moraity (i.e. MORAIC CODA) is too low ranked. Recall moreover that singleton codas can only be found word-finally since medially they do not occur.

---

one would exactly devise an OCP(weight) constraint. Perhaps, there is instead an upper bound on the moras a word can include, cf. similar restrictions in Bella Coola (Bagemihl 1998) or Yoruba (Ola 1995, 1997). For current purposes however, I will make use of the OCP.

Given the discussion above, Trukese renders itself a good candidate for a language with moraic onset geminates. For the most part, I have followed Muller's analysis, but this too presents a small bug. Because it locates  $\mu_*$  highly in the ranking, it implies that word-final mora deletion applies in the majority of cases. Thus, the prediction is that a  $/CV^\mu/$   $CVV$  root should actually surface as  $[C^\mu V^\mu]$   $C_i C_i V$ , because by doing so, both WDMIN and  $\mu_*$  can be satisfied. Davis' and Torretta's (1998: 121-122) solution to the problem employs the use of \*MISMATCH, which requires that moras in input-output correspondence must be realized on like segments, i.e. either on vowels or on consonants but not on vowels in the input and on consonants in the output and the reverse. This cannot extend as straightforwardly to the present analysis, because we do need a violation of \*MISMATCH for the case of  $/ta^\mu p^\mu a^\mu/ \rightarrow [ta^\mu p]$  since the mora of input  $/p/$  corresponds to one of the moras of the long  $[a]$  in the output. Since this is a minor technical point, which can be amended by modifying the constraints slightly, I set it aside for future discussion.

A final mystery that both analyses face is why a few bimoraic words present final mora deletion, without any subsequent lengthening<sup>28</sup>. For instance:

(39) *Final mora deletion in words of  $2\mu$  without lengthening*

Sfxed form		Unsfxed form		
a. kučču-k	'you suit'	kuč	'suit'	*kkuč, *kuuč
b. počče-y	'I hurry'	poč	'hurry'	*ppoč, *pooč

Davis and Torretta (1998) fail to mention the existence of such data, whereas Muller (1999) seems to suggest that she will provide an explanation of the facts, but this never comes. Additional data, information about the generality of the phenomenon and the

<sup>28</sup> I suggest that these data should be treated with caution. Muller only mentions these two examples, but I have been unable to find her example *kuč* in Goodenough's and Sugita's (1980) dictionary, which serves as the standard source for Trukese examples. A similar example exists with a relevant gloss, i.e. *kkuč* 'fit well', but this is no counter-example to the preceding discussion. Perhaps this example after all comes from Muller's own fieldwork data. As for *poč* (Goodenough and Sugita 1980: 286), this is presented as *počče-y* (suffixed) ~ *poča* (unsuffixed). It is accurate compared to (39), since it is assumed that all root forms underlyingly end in a vowel which is subsequently lost. Notice that the two examples that Muller uses however involve a final affricate. As Moira Yip (p.c.) observes, it would be possible to assume that in fact the affricate is a cluster, i.e.  $[tʃ]$  and as such the final consonant cannot carry a mora, but the previous one can, e.g.  $po^\mu t^\mu <ʃ>$ . While this proposal requires further in-depth study of Trukese, there are some reasons why we should not entertain it. First, Muller (1999: 401) states that consonant clusters are never allowed in Trukese. And second, a brief look in Goodenough and Sugita reveals that there are several other similar examples which do not involve final affricates such as: *kus* 'spurt out, get away', *nan* 'chatter', *tik* 'be sticking'. The thing all these share is that they are verbal forms (and sometimes adjectival too). As we have seen in fn. 17, verbs can be monomoraic in contrast to nouns which always have to be bimoraic.



structure of the Trukese lexicon are some of the things one would need to look into for a more well-informed exploration of this detail.

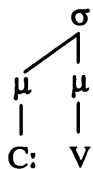
Meanwhile, the currently available Trukese data are entirely compatible and more straightforwardly accounted for by means of a ‘moraic onset’ analysis. What is more, with the grammar proposed, we may also predict exactly in which instances we should expect to find initial moraic geminates. These include: (i) when, regardless of how big the word is, there is only one geminate underlyingly and it is initial, e.g. /fféne/ → [ffén] ‘advice’, because of IDENT-μ >> \*MORAIC ONSET and (ii) in /C<sub>i</sub>C<sub>i</sub>VC<sub>j</sub>C<sub>j</sub>V/ unsuffixed inputs of exactly two syllables, e.g. /ppeppa/ → [ppɛp] ‘skip’ because of \*FINAL GEM >> \*MORAIC ONSET. An identical - or bigger - input to (ii), which nevertheless happens to be suffixed, will arise instead with the medial geminate because of OCP GEM<sub>{MORPH}</sub> >> \*MORAIC ONSET >> NO CODA.

## 6.4 Moraic onsets in medial geminates (‘Initial’ geminates word-medially)

### 6.4.1 Theoretical background

So far, I have argued that initial moraic geminates can be analysed as moraic onsets with the structure in (40). However, in §6.2 previously, I hinted at the idea that the same representation may also be possible word-medially. In other words, that we can get so-called ‘initial’ geminates word-medially too. In this part of the chapter, I explore this possibility both theoretically (in this section) as well as empirically (in the next section) considering data from Marshallese.

(40) *Moraic geminates as moraic onsets*

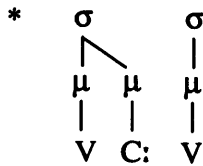


The possibility for the structure in (40) word-medially springs from an idea proposed by Ham (2001), who claims that the usual ‘flopped’ structure of a medial geminate, where the first half syllabifies in the coda and the second half in the onset of the next syllable is not necessitated by moraic theory but by syllable theory which prefers onsetful syllables. Moraic theory only asks that the geminate bears a mora. The fact that the

geminate is also doubly linked to a coda and an onset is the result of preferring more well-formed syllables, i.e. those beginning with an onset. To see this, compare (41a) where the geminate is wholly syllabified as a coda, thus rendering the next syllable onsetless, with (41b), where the mora of the geminate is still linked to the coda, but itself is split between the coda and the onset for syllabification purposes.

(41) *Geminates word-medially*

a. Avoidance of onsetless  $\sigma$ s



b. Preferred structure



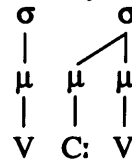
But there is an additional representation which simultaneously assigns a mora to the geminate and conforms to preferred syllabification. This possibility is not mentioned in Ham, but is made available in the current work which accepts the moraicity of the onset. Such representation is depicted in (40), where the geminate syllabifies wholly in the onset, but also carries a mora, which this time shows up in the onset rather than in the coda. The two possibilities, the flopped structure and the moraic onset geminate, are shown in (42). Crucially, in the former, it is the first syllable that becomes heavy, while in the latter, it is the second one.

(42) *[VC:V] sequence word-medially*

a. Avoidance of moraic onsets



b. Avoidance of syllables with codas



Languages may choose between (42a) and (42b). (42a) does better in having no moraic onsets at all, which entails that the constraints banning all types of moraic onsets - let us call this generically \*MORAIC ONSET - are highly-ranked. (42b) is better in avoiding extra codas. Assuming a language has \*MORAIC ONSET low-ranked and provided NOCODA >> \*MORAIC ONSET, the latter constraint will prefer a structure where codas are minimised by virtue of tautosyllabic assignment in onsets (as in (42b)). This ranking of NOCODA - if sufficiently low-ranked - does not entail that codas will be banned altogether (e.g. they should be allowed in non-geminate CC clusters), but that they will

be avoided wherever possible. Evidently, most languages rank \*MORAIC ONSET highly, consequently it is only natural that the majority of languages syllabify geminates heterosyllabically (42a) rather than tautosyllabically (42b). The latter is merely rarer, but by no means impossible. In the following section, I argue that Marshallese has geminates which syllabify as moraic onsets word-medially.

#### 6.4.2 Marshallese

The Ralik dialect of Marshallese (Micronesian) presents an interesting case of consonant doubling to form the distributive<sup>29</sup>. The distributive is an important category in Marshallese, which is very productive and covers a constellation of semantic distinctions (for discussion see Bender 1991). Most distributives are formed by consonant doubling and final syllable reduplication, e.g. *koto* → *kkototo* ‘windy → always be windy’. However, these two processes seem to be independent from one another since, as Bender notes (1991: 15), words which already present initial consonant doubling can form the distributive by having final copying, and those which already have final reduplication may instead present gemination only. The current discussion will thus address only the facts relevant to gemination<sup>30</sup>.

#### (43) *Marshallese distributive - Ralik dialect*<sup>31</sup>

<u>Root</u>	<u>Ralik</u>	<u>Gloss</u>
korap	yokkoraprap	‘gecko’
tumej	yuttumejmej	‘open eyes under water’
panuk	yeppanuknuk	‘pile up, gather’
bale	yebbalele	‘a type of fish’
diylah	yiddiylahlah	‘nail’
jekapen	yejjekapenpen	‘less than half full’
mede	yemmedede	‘young coconut meat’
nib	yinnibnib	‘preemptive’

<sup>29</sup> In the Ratak dialect of Marshallese, the reduplicant appears as a simple CV-copy, e.g. *betah* → *bebetah*, *diylah* → *didiylah*, etc. This is Hendricks’ (1999) interpretation. Other sources seem to suggest that there too gemination applies, but the geminate is split by an epenthetic vowel identical to the base vowel. The first view is simpler, but still it seems to me that the latter one projects the unity and the similarity of the process between the two dialects in a more direct way. Accounting for this formally however is more difficult.

<sup>30</sup> These data come from Abo et al. (1976). Next, I will build on data based on Zewen (1977). These two sources use different symbols for certain segments. In order that I avoid confusing the reader with presenting too many different characters, and given that Abo et al. (1976) is a dictionary, I had the luxury to pick out forms with segments whose representation is the same in both sources. With respect to data however from Zewen, I will introduce the symbols he uses.

<sup>31</sup> Following the practice of Hendricks, I present the reduplicated data not with a phonemic transcription, but with their spelling as presented in Abo et al. (1976). Apart from being a harmless simplification for current purposes, it is also preferred, as Abo et al. (1976) do not always give the phonemic transcription, so one would half the time need to guess what this would really be.

lokjak  
reja

yollokjakjak  
yerrejaja

'be busy with'  
'shave (from Engl)'

As Hendricks (1999: 36) observes: “(in Marshallese) the reduplicant (at least in most dialects) surfaces in onset position of a syllable”, a fact that makes this language relevant to our discussion. He nonetheless correctly claims that since traditionally it has been argued that onsets do not carry weight, no prosodic analysis that makes direct reference to them can be constructed. But of course, this is no longer necessary in a model along the lines proposed in this thesis. Since onsets can be weightful, we could express such facts by means of moraic onsets. I will argue that this is the case for reduplication in the Ralik dialect of Marshallese. In other words, one possible account of the reduplication facts is that the reduplicant is merely a mora and that the minimal way of realizing it is by geminating the consonant. This is entirely compatible with the ‘moraic onset’ proposal.

However, while virtually all sources agree that the above pattern is what generally goes on, the reduplicated form actually never surfaces exactly as suggested. Instead, an epenthetic *yV-* prefix precedes it. The quality of the vowel in *yV-* is the same as the base vowel, unless this is /a/ in which case the epenthetic vowel is /e/.

There are at least two issues which crop up at this point. First, how can we account for the presence of *yV-*? Isn’t this part of the reduplicated form? In other words, why don’t we simply get *reja* → *\*rrejaja* instead of *yerrejaja*? It seems that this can be straightforwardly accounted for if we claim that Marshallese bans initial geminates. This is not surprising given that most languages prefer medial to initial geminates. This could explain the presence of an epenthetic vowel word-initially. To account further for the onset preceding it, on top of that, we also need to add a requirement that syllables must have an onset. While this yields the prefixation of a *CV-* form to the reduplicant, it does not answer the question why *re-rejaja* is not chosen instead given that it fits the bill perfectly, and also happens to be the form that appears in the Ratak dialect.

At this point I can only speculate and will not offer a fully-fledged analysis, but the insight seems to be the following (cf. Hendricks 1999). Ralik requires absolutely minimal reduplication (Sloan 1988, Spaelti 1999), so it disfavours Ratak’s *CV-* reduplication and instead goes for *C-*reduplication. Moreover, the *yV-* prefix seems to have to be considered as somehow external to the reduplicated form. Its presence must be justified for phonological reasons only, i.e. avoiding an initial geminate and ensuring an onsetful syllable. Hendricks shares this idea, and claims that since *yV-* is epenthetic and corresponds to no input morphological material, it is not part of the morphological

word of the reduplicated form<sup>32</sup>, thus *yerrejaja* corresponds to *ye[r-rejaja]<sub>word</sub>*. In this word, the base *reja* is closer to the left edge of the word than it would be in the rival reduplicated form *[re-rejaja]<sub>word</sub>*, and thus it wins. But how exactly can we exclude that *yV-* is not part of the reduplicant since the latter is included in the input, i.e. /RED-*reja*/ and thus any part of its output realization must be included in the morphological word? Marshallese distributive reduplication thus definitely merits additional investigation, but at this point it seems that claiming that it involves consonant doubling is on the right track.

Having thus seen that geminates occur in Marshallese and are potentially moraic onsets, I will address another aspect of the language which leads to the same conclusion. This deals with the effects of geminates and stress. I will argue that stress provides evidence which suggests two things: i) Marshallese geminates are onset geminates, and ii) Marshallese geminates are moraic.

Since the stress data all come from Zewen (1977)<sup>33</sup>, let me first introduce the spelling conventions used in that work and adopted here to facilitate comparisons between the two. With respect to consonants, it suffices to mention that *b*=rounded bilabial plosive, i.e. something like *b<sup>w</sup>*, *m*=rounded bilabial nasal, i.e. something like *m<sup>w</sup>*, *ɳ*=*ɳ*, *l*=*l*, *r*=alveolar-coronal trill, *ɽ*=dental-apical trill.

The following represents the vowels of Marshallese (Zewen 1977: 31). Most vowels have short and long versions, thus:

(44) *Marshallese long and short vowels*

Short	Long
i [ɪ]	i: [i:]
e [e]	e: [e:]
a [a]	a: [a:]
u [ʊ]	u: [u:]
o [ɔ]	o: [o:]

Some vowels are always long or always short (45). Unlike Zewen, in the examples that follow I will make this clear, since I will be adding the length mark whenever appropriate.

<sup>32</sup> Van Oostendorp (2005) pursues this general idea and incorporates it in a larger framework entitled 'Coloured Containment'. As this theory is not yet totally spelled out, I will not attempt a re-analysis along those lines at present.

<sup>33</sup> As far as I have been able to tell, no other work of the ones I have consulted, i.e. Abo et al. (1976), Bender (1976), Bender (1991), Hendricks (1999) makes reference to Marshallese stress.

(45) *Marshallese long-only or short-only vowels*<sup>34</sup>

- a. Long only: low mid-front unrounded ä: [æ:]  
low back rounded ô: [o:]  
b. Short only: high central rounded ü [ʊ]  
high-mid central somewhat rounded probably ö [ø]  
low-mid central ɔ̄ [ə]

After this preamble we can move on to the stress facts (Zewen 1977: 40–41), where I use the acute accent to indicate stress. Stress is assigned within a trisyllabic window at the right edge. If the final three syllables are light, the antepenult gets stress. Codas - at least final ones - do not count for stress purposes (e.g. *ékajet* and not *\*ekajét*).

(46) *Trisyllabic words LLL: Antepenultimate stress*

ékajet	‘to judge’
núkileb	‘to have a big family’
jékaru	‘coconut syrup’
lákatib	‘to make angry’

If there is a heavy syllable, then this is stressed.

(47) *Trisyllabic words LHL or Disyllabic LH: Stress H*<sup>35</sup>

je.ú:rur	‘commotion, excitement’
je.ro.á:n	‘to waste’
körá:	‘woman’
jelä:	‘to know’
kijék	‘fire’

Penultimate stress appears on disyllabic words where both syllables are either heavy or light, i.e. the leftmost of the two is stressed (48).

(48) *Disyllabic words (either LL or HH): Penultimate stress*<sup>36</sup>

nébar	‘to praise’
má:ja:j	‘to be clear of underwood’

<sup>34</sup> Abo et al. (1976) have a somewhat different inventory particularly with respect to the vowels in (45b). I am not in the position to choose one over the other. The stress data I focus on hinge on vowel length, so if this is correctly presented, then slight differences in vowel quality are of no importance.

<sup>35</sup> Marshallese diphthongs are: ao, au, ai, ae, ei, öu. These act as long vowels. It is implied that the remaining vowel sequences are heterosyllabic.

<sup>36</sup> An exception is *kötó* which is not stressed on the penult as expected.

Default stress then is leftward (within the trisyllabic window), but it can shift so that it docks onto a heavy syllable. As we have mentioned already, coda consonants - or at least final codas - do not render a syllable heavy and thus do not attract stress. This proves crucial when we consider what happens in the case of medial geminates.

(49) *Geminate stress* (Zewen 1977: 27)

ji <b>bb</b> úŋ	‘morning’
(y)e <b>mm</b> án	‘good’
e <b>mm</b> ér	‘to be ripe’

I have managed to spot only four examples with geminate consonants, and stress clearly indicated. Three out of them (49) suggest tautosyllabic syllabification and contribution of a mora<sup>37</sup>. To show this, see what happens if syllabification were e.g. *ji**b**.buŋ*. Since codas do not contribute any weight, then both syllables should be light, in which case we would expect *ji**b**.buŋ*, because this is the leftmost syllable that can carry stress (cf. (48)). For the sake of the argument though, let us assume that only final codas do not contribute weight (maybe because they are extrametrical), whereas medial ones are moraic. Again, *ji**b**.buŋ* is predicted since it is the heaviest syllable (cf. (47)). But this is not the right result. The only representation that can give us the correct stress pattern is the following:

(50) ji <b>bb</b> úŋ	*ji <b>b</b> .buŋ
(y)e.m <sup>w</sup> m <sup>w</sup> án	*(y)ém <sup>w</sup> .m <sup>w</sup> an

The geminate is wholly syllabified in the onset of the syllable. Moreover, it contributes a mora, which renders the syllable heavy and thus stress-attracting. Unfortunately, no further data are available to shed more light on this issue. Given what we have available though, there is indication that Marshallese has onset moraic geminates which make their presence evident in both reduplication as well as stress.

One potential test for stress however is the following. Abo et al. (1976: xxxi) report that Marshallese commonly employs the prefix *ri-* with its various alternants to form Person nouns, e.g. *ruwa* ‘sailor’. Now, a number of person nouns have their first

<sup>37</sup> The exception is *jàrrik* ‘boy’ (Zewen 1977: 27) with stress on the first syllable. This suggests a heterosyllabic syllabification. There are plenty of examples where stress is not shown; some of these are: *jillüb* ‘sound produced by something that falls into the water’ (1977: 56), *killebleb* ‘to be very big, corpulent’ (1977: 59), *komman* ‘make’ (1977: 107), *ello:lo* ‘see’ (1977: 110).

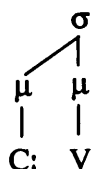
consonant doubled when this suffix is added, thus *tariaṇae* > *rūttariaṇae* ‘soldier’<sup>38</sup>. If stress assignment occurs as described below, then we can imagine the following situation: we take a base like *katta*, which we should expect to be stressed as *kattá* if the geminate is tautosyllabic, but *kátta* if it is heterosyllabic. Then we add the prefix *ri-*. Assuming that consonant doubling occurs, what we should get is something like *ri-kkatta*<sup>39</sup>. Given that the leftmost heaviest syllable receives stress, then depending on whether the geminate is tautosyllabic or heterosyllabic, we should expect *rikkátta* or *rikkatta* respectively. Anticipating relevant empirical testing, for the time being I merely observe that an analysis of geminates as moraic onsets in Marshallese seems plausible.

## 6.5 Remaining issues

### 6.5.1 Geminate weight and length

The preceding discussion points to a related issue: that of the connection between geminate length and weight. In this chapter I have used a representation of geminates particularly initially, but also medially, where the consonant in the onset position is simply moraic.

#### (51) Initial geminates



Notably, the geminate here does not link to any other prosodic structure, as is standardly the case in medial geminates that straddle a coda-onset boundary. One common view suggests that the latter representation has the advantage that it directly reflects the longer duration characteristic of geminates by means of the two links available. If this is true, then it poses a problem for the representation in (51), which lacks such double-linking.

Recall however that moraic theory argues that the distinction between geminates and singletons is not one between double- and single-linking. Rather it is the contrast

<sup>38</sup> For Abo et al. (1976), *n* = *nʷ*, *ū* = *uu*.

<sup>39</sup> Two geminates within a word are possible, e.g. *yibbiddikdik* ‘many crumbs, grains, morsels’ (Abo et al. 1976: 32).



between an *underlyingly* moraic consonant /C<sup>u</sup>/ versus a non-moraic one /C/. In other words, moraic theory on its own does not impose double-linking. As pointed out in §6.4.1, geminates in medial position, can achieve better syllabification by spanning the syllable boundary, since the geminate can be moraic in a coda position and provide an onset to the otherwise onsetless syllable that follows. In addition, the resulting double linking can also be interpreted as longer duration. However, we have argued that word-initially and word-finally no such double linking occurs, and yet, geminates are longer than singletons, although there is impressive variation in duration differences cross-linguistically. In fact, Ladefoged and Maddieson (1996: 91-92) report that, depending on the language, geminate stops can be anywhere between one and a half to three times longer than their singleton counterparts. So if double linking is not necessary, how can we contrast singletons from geminates? And how about the relationship between weight and duration in geminates?

To reconcile all these properties, I will follow the idea of Hubbard (1994) and its extension in Ham (2001) about the moraic primacy in the implementation of timing. The idea is that moras are allocated a minimum target duration, whose implementation in terms of timing takes precedence over any other segment-specific effects, such as place of articulation and voicing. If the main characteristic of geminates is to be moraic, then it is anticipated that the priority is to achieve the mora's minimum duration target, thus leaving smaller space for other durational segment-related differences.

Ham (2001) tested this hypothesis comparing the segment durational effects of voicing and place of articulation in four languages with geminates: Bernese, Madurese, Levantine Arabic and Hungarian. With respect to place of articulation, a strong tendency is that closure duration decreases the less anterior the point of oral constriction becomes (Ham 2001: 215 and references therein). Ham found that the place effect on closure duration is substantially larger in singletons than in geminates. Similarly, although voiced stops are generally shorter than their voiceless counterparts due to the aerodynamic difficulties of sustaining vocal fold vibration in the presence of a complete oral seal (Ham 2001: 217 and references therein), singletons exhibited more voice-conditioned variation in closure duration than the geminates.

Both these findings corroborate the anticipated prediction, namely that due to moraic primacy, geminates should present smaller durational effects on the segmental level compared to singletons. This is not however a prediction shared by a view that treats geminates not as moraic, but as inherently long and phonologically represents

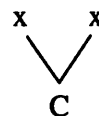
them through double-linking in terms of two timing slots or two root nodes (cf. Tranel 1991, Selkirk 1990 and others).

(52) *Geminate representations based on length*

Selkirk (1990): as two root nodes



Tranel (1991): as two timing slots



As Ham (2001: 223) observes, such models predict that segmental effects on duration should either be greater in geminates, if the magnitude of the effects directly correlates to the number of timing slots, or they should be the same between geminates and singletons, under the assumption that the effects are distributed across both the timing slots. At any rate, the empirically confirmed prediction that the effects are minimized, is not possible under this view.

Moreover, a two-slot geminate model seems to predict that all else being equal, geminates should have double the duration of singletons owing to the double association to timing structure. A moraic theory on the other hand only says that the geminate is underlyingly moraic leaving the phonetic implementation of phonological weight to language-specific considerations. This seems to be closer to reality, as Ham demonstrates empirically (for further discussion see Ham 2001: Chapter 7). All in all then, in the moraic theory of geminates put forward by Ham, the added duration of geminates observed in the phonetics, is the product of an underlying mora and not of double linking per se.

Focusing on phonological representations now, as we have seen (cf. §6.4.1), double-linking of a geminate in word-medial position achieves better syllable structure by providing an onset to the second syllable and assigning a mora to the coda of the first syllable, so  $VC_i^\mu.C_iV$  is preferred. This is what happens in a language that avoids moraic onsets (i.e. in most languages). A language that permits moraic onsets, e.g. Marshallese, will opt for the word-medial  $V.C_i^\mu V$  representation because it provides an even better syllabification by avoiding a coda at the expense of a moraic onset.

As a matter of fact, I argue that these are the only two representations a medial geminate can take. The other alternative  $VC_i^\mu.V$  (cf. (41a)) is excluded as it will always be harmonically bounded by  $VC_i^\mu.C_iV$  given that while both violate NOCODA equally, only the former also violates ONSET [to see this, in (53) compare (a) with (b) in each case]. If the constraints involved here are ONSET, NO CODA and \*MORAIC ONSET, then their factorial typology yields six grammars, half of which prefer the ‘flopped’  $VC_i^\mu.C_iV$  and half the moraic onset  $V.C_i^\mu V$ . Importantly,  $VC_i^\mu.V$  is never chosen. To illustrate, it only suffices to consider two tableaux.

(53) *Medial geminate factorial typology*

(i)	$/VC_i^\mu V/$	ONSET	*MORAIC ONSET	NOCODA
a.	$VC_i^\mu.C_iV$			*
b.	$VC_i^\mu.V$	*!		*
c.	$V.C_i^\mu V$		*!	
(ii)	$/VC_i^\mu V/$	ONSET	NOCODA	*MORAIC ONSET
a.	$VC_i^\mu.C_iV$		*!	
b.	$VC_i^\mu.V$	*!	*	
c.	$V.C_i^\mu V$			*

Here only the relative position of NOCODA and \*MORAIC ONSET differs. In the first case, moraic onsets are banned therefore the ‘flopped’ structure is selected. In the second case, avoidance of a coda is more important, therefore a moraic onset geminate is admitted. This ranking between these two constraints produces the variable representation of geminates. If NOCODA >> \*MORAIC ONSET, then  $V.C_i^\mu V$  is chosen, if \*MORAIC ONSET >> NOCODA, then  $VC_i^\mu.C_iV$  is favoured. For this reason, of the four remaining grammars in the factorial typology, the two with NOCODA top-ranked choose (53c) as the winner and the two with \*MORAIC ONSET top-ranked choose (53a)<sup>40</sup>.

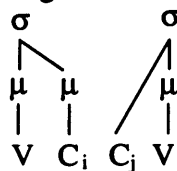
<sup>40</sup> If Anttila (1997) is right in deriving the frequency of outputs based on constraint rankings, then we would perhaps expect that given this factorial typology half of the time medial geminates should be syllabified tautosyllabically in a coda and the other half heterosyllabically in coda-onset configurations. This would be problematic, since cross-linguistically the overwhelming majority of languages prefer heterosyllabic syllabification. However, Anttila’s proposal is not unquestionable. The frequency effects crucially bear on the number of constraints assumed each time. To put it simply (and setting harmonic bounding aside), to get two possible outputs with a frequency of 50%-50%, then two constraints are needed; if the frequency changes to 66%-33%, three are required; if to 75%-25%, four and so on. To some extent then, the statistics can be manipulated depending on the constraints involved. If Anttila is wrong, then at this stage no concrete prediction can be made regarding the frequency of the two patterns with respect to geminates. If he is nonetheless correct, then we would need to identify how many languages there are with ‘initial’ medial geminates and then provide some fine-tuning of the factorial typology to capture this effect statistically too. I leave this matter open for further investigation.

Word-finally and word-initially though no such preferences apply and consequently we find single linking of a geminate to a mora yielding  $VC^{\mu}]#$  word-finally in languages like Hungarian and  $#[C^{\mu}:V$  word-initially in Pattani Malay and Trukese.

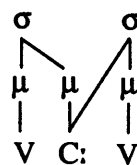
One question remains. If phonetics deals with the implementation of phonological weight, do we not run the risk of losing the surface contrast, e.g. between a singleton  $CVC^{\mu}$  where the coda acquires a mora through WBYP/MORAIC CODA and a geminate  $CVC^{\mu}{}^{\mu}$ , given that both are phonologically represented the same way? This problem essentially arises word-finally, because word-medially the preferred ‘flopped’ structure also happens to distinguish between the two, thus posing no threat (54).

(54) *Word-medial distinction between singletons and geminates*

a. Singleton  $CVC^{\mu}$



b. Geminate  $CVC^{\mu}{}^{\mu}$



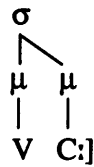
With regard to word-final geminates, Ham explains that in all the languages he examined (Bernese, Levantine Arabic and Hungarian), word-final singletons were treated as light (therefore it could be argued that WBYP was inapplicable), but geminates acted as heavy, a fact attributed to their inherent moraicity. This means that the contrast between singleton and geminates word-finally is once more retained in the guise of:

(55) *Word-final distinction between singletons and geminates*

a. singleton  $CVC]$



b. geminate  $CVC^{\mu}{}^{\mu}]$



Ham therefore makes the strong prediction that there will be no language which has both heavy singleton  $CVC^{\mu}$  and geminate  $CVC^{\mu}{}^{\mu}$  (although word-medially singletons can be moraic, given that the contrast with geminates is available through other means (cf. (54))). In the sample of languages he has investigated, this prediction holds true; nonetheless further empirical testing is in order.

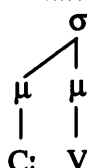
This now brings us to the last case: the initial position. Recall from the discussion in this chapter that in languages like Pattani Malay there is a contrast between singleton and geminate onsets, e.g. [ɟàlé] ‘road, path’ vs. [ɟ:ále] ‘to walk’. By assigning a mora to the initial onset of the second word, while leaving the corresponding one of the first word mora-less, the contrast can easily be maintained, as shown below.

(56) *Word-initial distinction between singletons and geminates*<sup>41</sup>

a. singleton CV



b. geminate C<sup>μ</sup>V



At this point, one final remark is in order. CVC<sup>μ</sup> singleton codas acquire weight on the surface - usually due to WBYP/MORAIC CODA - thus they illustrate the coerced type of coda weight (Morén 1999/2001). CVC:<sup>μ</sup> geminate ones on the other hand are underlyingly moraic, hence they represent an instance of distinctive weight. As these types of codas can co-exist in languages (e.g. Italian), one wonders whether a similar situation may arise in onset weight. In other words, should we expect to find languages where coerced onset weight of e.g. the Pirahã type and distinctive onset weight of e.g. the Pattani Malay type are simultaneously present?

I suggest that this is impossible. First observe that in the case of codas, the contrast between coerced and distinctive ones is represented either in syllabification, that is geminates have the ‘flopped’ structure medially (54) or some sort of extrametricality of singletons word-finally (55). For onsets, this is not possible. Recall that apart from geminate moraic onsets (distinctive weight), there are also moraic onsets which are the product of coerced weight, i.e. weight imposed by MORAIC ONSET as in Pirahã (Chapter 2) or WDMIN as in Bella Coola (Chapter 4). However, the syllabification of coerced and distinctive moraic onsets is the same and simply contrasts with non-moraic onsets, i.e. (56b) vs. (56a).

Now suppose we find a language that contrasts [pa] with [p<sup>μ</sup>a] and [ba] with [b<sup>μ</sup>a]. The first thing that would come to mind given this distribution is that this language has distinctive onset weight. If it were to also possess coerced onset weight,

<sup>41</sup> The same contrast is also available word-medially for cases like Marshallese with ‘initial’ geminates.

then only contrasts of the type [p<sup>μ</sup>a] and [ba] due to voicing should be permitted (see Chapter 1 and 2). And yet [pa] and [b<sup>μ</sup>a] are also allowed, which indicates that coerced weight is inapplicable here. Similarly, if a language only has a [p<sup>μ</sup>a] and [ba] contrast, then it would need to be of the coerced type, since under distinctive weight the lack of [pa] and [b<sup>μ</sup>a] is unexplained. For such cases then, moraic onsets would either take the form of coerced or distinctive weight, but not of both.

Seeing this point from a different angle, note that if a language is found to have [p<sup>μ</sup>a] and [b<sup>μ</sup>a], it does not necessarily mean that it provides evidence for distinctive weight. Recall that in coerced onset weight the markedness hierarchy that holds is: \*μ/ONS/[voi] >> \*μ/ONS which always prefers moraic onsets that lack specification of [voice] over those that possess it. In addition, the constraint MORAIC ONSET is interleaved among these constraints assigning moraicity to certain types of onsets. Although in Chapter 2 we had clear evidence for the ranking: \*μ/ONS/[voi] >> MORAIC >> \*μ/ONS, which only admits moraic onsets without [voice], another pattern we could also have is: MORAIC >> \*μ/ONS/[voi] >> \*μ/ONS. This latter ranking assigns onset moraicity to all types of onsets, effectively generating [p<sup>μ</sup>a] and [b<sup>μ</sup>a] on the surface.

This is then a potential case for distinctive weight, but as we have just seen, for coerced weight too, so could it be an instance where both types co-exist? The answer is no, because they make mutually exclusive predictions. If it were distinctive weight, empirically we should expect to find contrasts between bimoraic [p<sup>μ</sup>a] and [b<sup>μ</sup>a] with monomoraic [pa], [ba] and [a] (cf. Pattani Malay). If it were coerced weight however, then the available contrasts should be between bimoraic [p<sup>μ</sup>a] and [b<sup>μ</sup>a] with monomoraic [a] only (cf. Bella Coola Word Minimality). At any rate, the empirical data would tell us that it should be one or the other, but not both. Given the data available in this thesis, this result seems to be correct, since no single language possesses both moraic onset geminates and moraic onset singletons.

In the next section, I will briefly consider one more case, that of alleged ‘weightless geminates’ and argue that since these lack weight, then they cannot be real geminates. A different representation is instead appropriate.

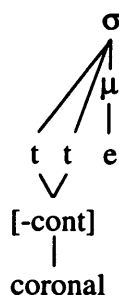
### 6.5.2 ‘Weightless geminates’

In spite of the results above, it has been claimed in the literature that languages with non-moraic geminates also exist, e.g. Malayalam, Selkup and Tübatulabal (Tranel

1991). These so-called 'weightless geminates' cannot be represented as suggested above, given that they lack a mora. They also seem to be in direct contrast with the proposal of moraic theory that geminate consonants are underlyingly moraic, and thus should be consistently weightful.

However, in the current view (see also Ham 2001, Hayes 1989), these are not really geminates, since geminates are by definition underlyingly moraic. Rather, they are simply doubled consonants with two root nodes (cf. Selkirk 1990), which can receive the following representation (also used in Hayes for ‘fake’ geminates which arise under morphological concatenation). These involve a complex onset made of two identical consonantal root nodes.

(57) *Representation of fake geminates and of doubled consonants*



That this representation seems well-suited for at least some of the languages Tranel discusses becomes evident when we consider, for instance, Malayalam. K. P Mohanan (1986: 73-74) and T. Mohanan (1989) argue that Malayalam bans codas and ‘weightless geminates’ are tautosyllabic in the onset position. Mohanan (1989) is more specific in claiming that the language actually allows codas during initial syllabification, but bans them in later syllabification. It is in the latter stage where this type of geminates are tautosyllabic, as evinced by native speaker intuitions, language games and word-level restrictions. Given the representation in (57), geminates in Malayalam can thus be argued to be an instance of doubled consonants.

Note that this representation is now extended beyond doubled consonants generated through morphological concatenation to identical consonants brought next to each other by a phonological accident. In all likelihood, this would then be the correct representation for alleged geminates in certain languages to be examined in the next chapter, such as *Tukang Besi* (Donohue 1999) and *Berawan* (García-Bellido and Clayre 1997). Since these consonants provide no evidence for moraicity and seem to require syllabication wholly in the onset, they are best represented as doubled consonants.

A final point of interest is that since geminates and doubled consonants are different species, one may expect to find them co-occur in the same language. This is what Ham (2001) argues for the case of Bernese, since he treats word-medial and word-final geminates as real geminates, but word-initial ones as doubled consonants resulting from morphological concatenation.

## 6.6 Conclusion

This chapter has dealt with distinctive onset weight, as it appears in both initial and medial geminates. Syllabification and moraification of an initial moraic geminate wholly in the onset resolves one of the longstanding problems of standard moraic theory, which suffers due to the inherent inability to represent a geminate in the familiar coda-onset structure in word-initial position. Admitting moraic onsets settles this problem since the geminate can now be simultaneously syllabified and moraified without any ad hoc assumptions that either leave part of it unlinked to any structure (Davis 1999) or unsyllabified and linked to higher prosodic structure (Curtis 2003). Implementation of this idea in Pattani Malay and Trukese successfully accounts for the facts at hand.

The next logical step has been to extend this proposal to word-medial geminates. As Ham (2001) observes, moraic theory only asks that a geminate carries a mora. Its split syllabification as argued for by the standard moraic model (Hayes 1989) is a by-product of syllable well-formedness, but not necessitated by the weight theory itself. Ham does not accept moraic onsets, but in the current work, where these are put forward, it is logically possible and in fact attested in a language like Marshallese to have a /CVC<sub>i</sub>C<sub>i</sub>V/ form (where C<sub>i</sub>C<sub>i</sub> = C: = geminate), syllabify as CV.C:V instead of the more familiar CVC<sub>i</sub>.C<sub>i</sub>V. The ranking NOCODA >> \*MORAIC ONSET allows for this possibility. The reverse produces the opposite result, i.e. the case where medial geminates syllabify as coda-onset instead of as exclusively (morai) onsets. The addition of moraic onset geminates, which had not been discussed in previous works, such as Ham (2001), once more highlights (cf. Ham 2001: 224) the asymmetrical representation of geminates in different word-positions.



## **Chapter 7**

### **Ambiguous and misanalysed onset-sensitive languages**

#### **7.1 Introduction**

In the previous chapters, I have been arguing for the existence of onset moraicity by investigating onset effects on weight. Several case studies had been explored to this end. In this chapter, this effort is continued, but takes a slightly different route. In particular, I examine a range of languages that fall within two broad categories.

The first refers to languages whose data lend themselves to a moraic onset analysis, but no conclusive evidence is available. In particular, these are cases where data are either too sparse or too ambiguous to reach a safe conclusion. The second type comprises languages which have in the past received analyses that made explicit or implicit use of moraic onsets. I attempt to show that this is the wrong analysis for these languages, or that at the very least, the evidence currently available does not strongly support such an analysis.

Thus the aim of the chapter is twofold; first, it highlights the fact that there are several languages which suggest but do not yet establish their use of moraic onsets. By addressing these languages and pointing out the questions these pose, it is anticipated that future documentation and investigation should be able to clarify their status. At the same time, this study shows what kind of traits moraic onset languages may have and therefore opens up the possibility of identifying additional languages that may fall within the moraic onset realm.

However, I contend that the reverse is equally important, that is, by examining languages which have been falsely or less convincingly argued to present themselves as candidates for onset moraicity, we learn more about the factors that can cause confusion and lead us to the wrong conclusions regarding their moraicity status.

The chapter is divided into two large chunks, namely sections 2 and 3. The former investigates several cases of languages whose data may support onset moraicity, but only if certain conditions are met. In most cases, further empirical confirmation is required to corroborate any relevant claim. To this end, numerous weight-sensitive effects are considered such as gemination, reduplication, CL, word minimality and

metrics. Section 3 on the other hand explores cases which have formerly been argued to illustrate instances of onset moraicity. I show that such analyses are misguided and their conclusion is wrong, since other data suggest otherwise. The languages that are examined are Berawan (with respect to geminates), and Bislama and Nankina (with respect to complex onsets and stress).

## 7.2 Languages suggestive of onset moraicity subject to satisfaction of certain conditions

This section consists of six sub-sections, each of which explores one or more languages. Section 7.2.1 looks at geminates with respect to reduplication and examines Bellonese (§7.2.1.1). Section 7.2.2 focuses on CL that is caused by onset deletion as in Onondaga (§7.2.2.1) and Proto-Austronesian (§7.2.2.2), but also, more remarkably, on CL that leads to the creation of a moraic onset as in Trique (§7.2.2.3). Section 7.2.3 looks at word minimality from a slightly different point of view as attested in Damin (§7.2.3.1). §7.2.4 briefly considers the metrics of Luganda court songs, whereas section 7.2.5 investigates a few more languages where onset moraicity is more contentious. These include Supyire geminates, vowel elision and CL (§7.2.5.1), Yoruba word-size effects (§7.2.5.2) and Tukang Besi geminates and stress (§7.2.5.3).

### 7.2.1 Geminates and reduplication

#### 7.2.1.1 Bellonese (Elbert 1988)

Bellonese (Melanesian, Solomon Islands) is another potential case showing reduplication by means of moraic onset gemination (cf. §6.4.2 Marshallese). Bellonese syllables are of the (C)V(V) type, thus no codas are permitted. Interestingly, all consonants can geminate in fast speech, whenever reduplication is involved [N.B the reduplicant is shown underlined].

#### (1) Bellonese long consonants in reduplication (Elbert 1988: 17-18)

Slow speech	Fast speech	Gloss
<u>b</u> abange	<u>b</u> bange	'to play'
<u>b</u> ebete	<u>b</u> bete	'to untie'
<u>h</u> ahatu	<u>h</u> hatu	'to fold'
<u>k</u> akata	<u>k</u> kata	'to laugh'
<u>l</u> olongi	<u>l</u> longi	'weak'

<u>m</u> omoe	<u>m</u> moe	'to lie down'
<u>n</u> anamu	<u>n</u> namu	'to smell badly'
<u>n</u> gangaha	<u>n</u> gngaha	'to measure'
<u>p</u> ipiki	<u>p</u> piki	'to keep'
<u>s</u> asaka	<u>s</u> saka	'to beg'
<u>t</u> atau	<u>t</u> tau	'to tattoo'

Elbert seems to attribute this phenomenon to the slow vs. fast speech distinction and thus implies that the geminated form is somehow generated after the deletion of the vowel of the CV reduplicant. To corroborate that, he claims that similar losses occur word-medially in some non-reduplicated words.

Nonetheless, I contend that a more interesting explanation of the facts is the following; the template for reduplication is merely a mora. Speakers can achieve this in two ways. They can either produce a CV syllable or they can geminate the first base consonant. While a double consonant is ideal in the sense that it is really the most minimal form one can get for mora reduplication, it introduces geminates which are quite marked. A CV-reduplicant avoids this problem at the expense of constructing a slightly larger form. Two pressures are then in conflict producing variation; on the one hand the preference for minimum reduplicant size and on the other the avoidance of geminates. Depending on which of the two takes precedence over the other, the above results obtain.

There is another reason why this analysis seems to be on the right track. Elbert notes that this kind of pattern *systematically* arises only in reduplication. A very limited number of similar cases emerge outside reduplication, e.g. *ghaghaghaba* ~ *ghghaghaba* 'a plantain'<sup>1</sup>. In fact, Elbert offers only two more examples of the same type. Had the pattern been described in (1) were purely related to the slow~fast speech difference, then why would it regularly and systematically only arise in reduplication? And why would it always and only involve deletion of the first vowel given that word-medial vowel deletion also occurs, e.g. *ghaghaghasa* ~ *ghghaghasa* 'a limpet'? The truth is that vowel deletion happens extremely marginally outside reduplication and seems random. The reduplicated pattern however is suggestive of a generalisation which can much more adequately be attributed to initial-consonant gemination that involves construction of a moraic onset - a configuration also supported by the prohibition against codas.

Bellonese then renders itself a compelling case of reduplication by means of initial consonant moraic gemination.

---

<sup>1</sup> Note however that despite Elbert's claim, this still looks like reduplication.

## 7.2.2 CL

### 7.2.2.1 Onondaga (Michelson 1988)

Onondaga resembles Samothraki Greek, which was previously discussed, in that deletion of /r/ results in lengthening. Facts are quite complex, so I will present a brief overview of the major points. Onondaga is one of the five Lake-Iroquoian languages, the other four being Seneca, Cayuga, Oneida and Mohawk. Onondaga and Seneca lost the phoneme /r/ of the Proto-Lake-Iroquoian (PLI) in all environments, whereas Cayuga lost it intervocalically and after a postvocalic laryngeal. The discussion here centres only around Onondaga, unless reference to the other languages is needed.

Intervocalically, *r* became *w* after round vowels (2a) and *y* after *i* (2b). In all other occasions *r* was lost (2c). The latter case resulted in a sequence of vowels, but no lengthening appeared<sup>2</sup>.

- |     |    |                  |              |                 |
|-----|----|------------------|--------------|-----------------|
| (2) | a. | PLI *yothó:re?   | 'it is cold' | Onond othó:we?  |
|     | b. | PLI *owí:ra?     | 'baby'       | Onond owí:yæ?   |
|     | c. | PLI *wa?knóhare? | 'I wash'     | Onond wa?koháe? |

More interesting is the pre- and post-consonantal *r* loss. In both cases, *r* deleted and caused lengthening of the vowel before it (pre-consonantal *r* e.g. PLI: \*katórye?s 'I'm breathing, Onond: katórye?s) or after it (post-consonantal). The latter is where we focus our attention.

#### (3) Post-consonantal /r/-deletion and lengthening in Onondaga

- |                  |              |                  |
|------------------|--------------|------------------|
| PLI *yotshí?kre? | 'clouds'     | Onond otshí?ke:? |
| PLI *ó?kra?      | 'snow flake' | Onond ó?kæ:?     |

Lengthening did not normally occur after \**hr* or \**?r*. Michelson states that the derived vowels are truly long and tautosyllabic since they are counted as one vowel for stress purposes. Seneca on the other hand presented *r*-deletion in the same environment without lengthening. Citing Woodbury (1981), Michelson proposes to explain the Onondaga pattern based on two facts. First, in dictionaries of the seventeenth and mid-eighteenth centuries, \**Cr* is frequently represented as *Cer*. Second, Mohawk regularly

<sup>2</sup> If the vowel following /r/ is back, then it is fronted, e.g. a → æ as in (2b).

inserts an epenthetic *e* to break up \**Cr* clusters. One could then assume that Onondaga at some point inserted an epenthetic *e* too, which after *r*-deletion would coalesce with the following vowel leading to a fused long vowel, i.e. \**CrV* > \**CerV* > \**CeV* > \**CVV* > \**CV*∶.

While this is a possible explanation of the facts, I do not think it is the only one. Onondaga's *r*-deletion patterns seem to be consistent with what happens in SamG, i.e. lengthening of a vowel after cluster simplification, but not when the *r* deletes word-medially<sup>3</sup>. With regard to the information the dictionaries offer, numerous questions can be raised, but I think the most pertinent one is why only some of the \**Cr* forms are represented as \**Cer*. Wouldn't one expect all the forms to show up with this pattern had it been so pervasive? More importantly, other than the dictionaries, there does not seem to be any conclusive empirical evidence that indeed there was *e*-epenthesis in the \**CrV* context. The fact that it occurs in Mohawk does not imply that it happens in Onondaga. What is more, Mohawk has extensive *e*-epenthesis, which is in direct contrast to the distribution of *e*-insertion in Onondaga. As a matter of fact, Michelson herself (1988: 146-7) describes just a single case of *e*-epenthesis, namely between an extrasyllabic non-laryngeal consonant and a following consonant (4).

(4) *Onondaga e-epenthesis*

k-theʔt-haʔ > k <u>e</u> théʔthaʔ	'I'm pounding'
k-ʔny-a-ʔke-h > k <u>e</u> ʔnyáʔkeh	'my hand'

I contend that all these facts show that this issue has not been resolved and should at the very least make us sceptic towards the \**Cer* approach. A compensatory lengthening approach is available, especially given the resemblance of the facts to the SamG data. This matter then remains open for further investigation.

### 7.2.2.2 Proto Austronesian (Zewen 1977)

According to Zewen (1977: 9-10, 36), initial Proto-Austronesian (PAN) consonants that were lost in Marshallese (MAR) have generally produced a compensatory lengthening effect as exemplified below.

<sup>3</sup> Onondaga differs from Sam. Greek in also presenting deletion of /r/ from the coda with subsequent lengthening. With respect to Onondaga initial /r/, Michelson (1988) reports that *stem*-initial /r/ alternates with y/Ø, but at least in the examples provided, this does not coincide with the word-initial position.

(5) Proto-Austronesian initial consonant loss and CL in Marshallese

PAN	MAR	Gloss
bitun	i:ju	'star'
binih	i:ne	'seed'
yaput'	ä:ut	'to wrap around'
pukət	o:k	'drag net'
puna	u:n	'origin'
hat'ap	a:t	'smoke'

However deletion in a cluster did not lead to CL, as in PAN *\*mpənəd* > MAR *bon* 'obstructed' [NB: *b* stands for a rounded bilabial stop, *t'* for an unreleased dental stop].

This set of data is of course promising, but additional empirical confirmation is required, particularly as Zewen does not explain what other processes may be involved in the evolution from PAN to MAR. It is worthwhile mentioning that one of the examples is interesting for an additional reason. Based on the Samothraki Greek 'compensatory lengthening after onset loss' data (Chapter 5), we have claimed that there is the expectation that a language may show lengthening after initial onset loss, but not after medial one, because the latter is easily blocked by the cross-linguistic tendency against super-long vocalic hiatus. This seems to be reflected in the example *yaput'* > MAR *ä:ut*, because the Marshallese form presents deletion of both the initial as well as the medial onset. Lengthening however occurs only after the former's loss.

### 7.2.2.3 Trique (Hollenbach 1977)

The description of Trique by Hollenbach presents an interesting case of CL, which is different from what we have seen before. While in all the previous cases, including Samothraki Greek, CL was seen as the result of onset loss which led to lengthening of the vowel following the onset, Trique shows a different pattern. Essentially here, the vowel is lost or shortened and the preceding consonant lengthens as a response to that<sup>4</sup>. This is exemplified below.

Trique is a tonal language, which also shows stress, consistently on the final syllable. This syllable is special in being able to license numerous tonal, consonantal and vocalic contrasts (see Hollenbach 1977 for details or Yip 2002: 234 for a summary). One of the contrasts in the San Juan Copala dialect of Trique (henceforth SJC), involves the presence of fortis stops and fortis sibilants in the onset of the final syllable and

<sup>4</sup> Something similar happens in Supyire vowel elision (§7.2.5.1), but facts there are more contentious.

nowhere else. The fortis consonants are voiceless and unaspirated (Hollenbach 1977: 36-37). Phrasal stress is marked both by the intensity on the ultima of the word receiving it and by lengthening of open long vowels or shortening of short vowels. “The latter is accompanied by *compensatory lengthening of an immediately preceding fortis stop, fortis sibilant or resonant* (Hollenbach 1977: 49; emphasis added mine)”.

If this is accurate, then it seems to be the case that the short vowels are now reduced, and their mora is inherited by the preceding onset consonant. Given that it is the stressed syllable which gets affected and that Copala Trique lacks codas (with the exception of word-final coda, which has to be a laryngeal (Hollenbach 1977: 36)), it seems reasonable that the newly created geminate syllabifies wholly in the onset and is rendered a moraic onset (cf. Chapter 6).

An additional desirable aspect of this is that only the fortis/voiceless consonants undergo this process. While lenis/voiced consonants can also surface as onsets of final syllables, cf. *nata?* ‘to explain’ vs. *roda?* ‘muller’, they do not undergo any lengthening. This is consistent with the idea pursued in this thesis that if only one set of consonants emerges as moraic onsets, then it will be the voiceless one to the exclusion of the voiced counterpart.

One further effect is also of importance. Hollenbach (see above) also states that resonants, i.e. /m n l y w/ undergo this lengthening too, which means that they behave like the voiceless consonants do. As a result, SJC Trique sonorants have to lack the [voice] feature. This property brings them on a par with Karo sonorants (see §2.4.3.1).

A similar instance of CL appears in the San Andrés Chicahuaxtla (SAC) dialect of Trique. SAC differs from SJC in having both fortis resonants /m: n: l: y: w:/ which only appear in monosyllables and lenis sonorants /m n l y w/. The characteristic of fortis resonants is that they are quite long (Hollenbach 1977: 50). Interestingly, SAC fortis resonants developed mainly by a lengthening of simple resonants to compensate for the loss of a penult (Longacre 1957: 18)<sup>5</sup>.

- (6) *SAC Trique fortis resonants* (Hollenbach 1977: 58; a dot underneath a vowel indicates it is short)

SJC	yume	SAC	m:ĩ	‘sweet potato’
SJC	yana	SAC	n:a	‘loft’

<sup>5</sup> Note that the Trique CL facts would require some modification to the CL theory presented in Chapter 5 with regard to assumptions about root node preservation in reduced vowels (SJC) and the addition of only one mora even when two segments delete (apparently due to limitations of the number of moras per syllable) as in SAC.

SJC    yuwe                      SAC    w:e                      ‘century plant’

If CL is either about mora preservation, or position preservation through a mora, then the fortis sonorant will carry a mora in SAC. And if initial moraic geminates are best described as weightful onsets, then this is another case of CL that leads to a moraic onset.

### 7.2.3 Word Minimality

#### 7.2.3.1 Damin (Hale 1981 (H<sup>6</sup>), Hale and Nash 1997 (H&N))

Damin is a now extinct initiation language used by the Lardil (Mornington Island, Gulf of Carpentaria, Australia). It is phonologically interesting for many reasons. It is the only language outside Africa that possesses clicks and shows rearticulation of some consonants. Damin is interesting in another respect too. According to H&N, it is also the case that the language has a word minimum which can be stated as CVV or CCV, while CVC minimal words seem to be banned, although Damin allows closed syllables with /n/ or /rr/ as a coda [/rr/ is an apico-alveolar flap]. There is however a preference for syllables to be open (H&N: 254). While I will not attempt to justify the consonantal inventory that Hale and Nash present, I will repeat it here for expository reasons.

#### (7) Damin consonantal inventory

	bilabial	lamino- dental	apico- alveolar	apico- domal	lamino- alveolar	dorso- velar
stops	b	th	d		j	k
nasals			n		ny	ng
flap			rr			
glides	w				y	
fricative	f					
	pf				j2	
ejectives	p'					k'
nasals	m!		n!	(m!)		
		nh!2	n!2			ng*
lateral			l*			
trill	pr					
	pr2					

The interested reader is referred to H&N: 252 for the detailed IPA transcriptions of these sounds. For current purposes it suffices to say that the symbol ‘2’ denotes

<sup>6</sup> Many thanks to David Nash and to ASEDA, AIATSIS for making document 0028 available to me.



rearticulation, ‘\*’ indicates that the sound is voiceless and ‘!’ signifies a standard click at the place of articulation denoted by the previous consonantal symbol. All clicks are nasalized. Intervocalic clusters are limited, but there are some. Onsets include singleton consonants, rearticulated consonants or certain clusters. In particular, the latter involve: a) the bilabial /f/ or /p’/ followed by /ny/ or /ng/, b) the bilabial /f/ or /pr2/ followed by /y/, c) the cluster /thrr/.

On a highly speculative note, Hale and Nash (1997), following a suggestion by Morris Halle, propose that in fact all complex onsets in Damin, including the rearticulated ones, have to be analysed as “the double association to onset of a single segment”. In other words, they are treated as geminate consonants. This is consistent with the fact that minimal words contain either a long vowel or are of the form CCV<sup>7</sup>, where CC corresponds to consonant clusters or rearticulated consonants. Both types of words are treated as bimoraic. Hence the onset is allowed to contribute a mora, a possibility that Hale and Nash (1997) themselves acknowledge. Some examples follow:

(8) Damin minimal words (H&N and H)<sup>8</sup>

CCV: a) rearticulated consonants

n!2a	‘grandmother, paternal’ (H)
n!2u	‘liquid’ (H&N)
nh!2u	‘become hot’ (H)

b) complex onsets

fngu	‘be bewitched’ (H)
fnyi	‘calf of leg or tail’ (H)
thrru	‘woman’ (H)

CVV: jii	‘grandchild, daughter’s child’ (H)
thii	‘shark (general term)’ (H)
kaa	‘now’ (H&N)

Note that some - superficially - CV words are allowed too, but these always involve one of the following consonants: /l\*/, /ng\*/ or /k’/, such as, l\*i ‘good, well’ (H), ng\*i ‘tobacco’ (H). This detail could destroy the bimoraicity minimality generalisation, but the solution seems to lie in the fact that these consonants too seem to be acting actually as rearticulated. /l\*/ has two sequentially opening airstream mechanisms, i.e. an

<sup>7</sup> The examples of rearticulated consonants provided in the sources only include rearticulated clicks. While there are words starting with /pr2/ and /j2/, these are followed by VV instead, e.g. pr2yuu and j2uu (H). They were the only examples of this kind I could spot.

<sup>8</sup> Document 0028 (Hale 1981) comprises a Lardil-Damin vocabulary and a Lardil-English one. The former provides no English glosses for words. I have provided the glosses indirectly by consulting the Lardil-English vocabulary. I usually provide a single gloss, since for some words multiple correspondents are applicable, e.g. n!2a also means ‘brother in law’.

ingressive voiceless lateral with egressive glottalic release, while /k'/ and /ng\*/ have 'double effort' in the airstream with /k'/ being an ejective and /ng\*/ having an extra pulmonic pressure (H&N: 257-258).

While the Damin data seem intriguing, more would need to be understood about the nature of the consonants involved and the possible clusters, as well as of the properties of rearticulation. However, chances to do so adequately are slim given that the language is extinct, therefore no additional data collection is possible.

## 7.2.4 Metrics

### 7.2.4.1 Luganda court songs (Fabb 1997, Katamba and Cooke 1987)

Fabb (1997) discusses Luganda court songs, which are part of the court music of the Buganda kings of Uganda, and argues that these songs are metrical (for more details see Fabb 1997: 100-102).

The basic line consists of 36 pulses divided into six 6-pulse constituents matched by accompanying handclaps. The music pulses match with moras in the line. For our purposes, it is interesting to examine the mora-counting algorithm. All the following configurations count as two moras: i) a long vowel, ii) a vowel preceded by a glide, e.g. *kwa* and iii) a short vowel followed by two consonants (usually a geminate or a nasal+obstruent sequence). The following line illustrates [N.B: • = pulse, ◡ = implicitly accented pulse accompanied by handclap].

#### (9) line from the song of *Ssematimba and Kikwabanga*<sup>9</sup>

Call part										Response part									
•	◡	◡	◡	◡	•	◡	•	◡	◡	◡◡◡	◡	◡	•	◡	◡	◡	◡	◡	◡
													U						
μ	μμ	μμ	μμ	μμ	μ	μμ	μ	μμ	μ	μ	μμ	μ	μ	μμ	μ	μμ	μ	μμ	μ
A-baa	- lia	- ban	- gi	nsi	ga	- dde	bw'o	- mu	La	- ba	Sse	- ma	ti	- mba	ne	Ki	- kwa	- ba	- nga
we	were	many	I	am	left	alone	Look	at	Ssematimba	and	Kikwabanga								

It is evident that the line comprises two half-lines: the call part and the response part. A few pulses may remain unmatched if the half-line is short. Notably, unmatched pulses

<sup>9</sup> The syllable [ma] in *Ssematimba* is monomoraic but matched to two pulses, as a result of lengthening in that position. The monosyllabic word [ne] that follows is treated as bimoraic, because the following consonant /k/ in *Kikwabanga* is pronounced [tʃ] and acts like the geminates or nasal+stop sequences (see below in text) rendering the preceding syllable heavy.

only occur at the end of a half-line, but never within it. The reader is referred to Fabb (1997) for more details on the analysis of this metrical template, but what matters for our purposes is that syllables with complex onsets like *bw'o* or *kwa* count as heavy for metrical purposes. One of course could argue that such syllables are truly *bw'o:* or *kwa:*, but I do not consider this likely, since Fabb is cautious in being explicit for all other vowels involved by marking them as long or short<sup>10</sup>. I thus take this to really indicate a short vowel preceded by a glide, but being able to act as bimoraic. Nonetheless, note that it is also possible that the glides here are what Smith (2003) calls 'nuclear onglides', i.e. glides that are actually dominated by the nucleus node instead of the syllable node. As such, they could contribute to the weight of the nucleus and consequently not count for onset moraicity. To entertain this possibility, further research would be required on the status of Luganda glides.

However, even if this is so, that proposal cannot be extended to the following observation. Syllables with a short vowel followed by a nasal+obstruent stop or with a geminate, e.g. *ti* in *ti-mba* or *ga* in *ga-dde* are considered bimoraic, perhaps due to resyllabification, i.e. *tim-ba* and *gad-de* from the onset to the coda of the previous syllable. But this may not be necessary if we assume, as Fabb does above, that complex onsets may contribute moras for metrical purposes. Although the correspondence between pulses and moras may need to be slightly modified, the number of moras is not affected, nor is the overall pattern.

Luganda court songs then provide a potential case where onset moraicity can be recognised in metrical structure.

## 7.2.5 Remaining cases

### 7.2.5.1 Supyire (Carlson 1990, Hajek and Goedemans in prep.)

In this section I describe a process of vowel elision in Supyire spoken in Mali (Carlson 1990), which according to an unpublished manuscript by Goedemans and Hajek (in prep.) could be treated as involving onset weight. Throughout I will not mark tone.

The facts of interest are the following (Carlson 1990: 67-68). Unstressed high vowels /i/ and /u/ delete when followed by a resonant consonant /l/ and /n/ (but

<sup>10</sup> Bear in mind that Fabb's interpretation of the facts conflicts with that of Clements (1986), where *kwa* is actually [kwa:] from underlying /ku-a/. This undergoes glide formation and lengthening of the following vowel. Similarly, *timba* is actually [ti:mba] from underlying /tim-ba/. Prenasalization follows with subsequent lengthening of the previous vowel.

apparently not when the resonant is a different nasal or an approximant /w/ or /y/) and preceded by either a non-coronal stop (/p/, /b/, /k/) or a non-back fricative (/f/, /v/, /s/, /z/)<sup>11</sup>. There are two options for the resulting surface form, either the one with consonant gemination, i.e.  $C_1V_1C_2V_2 \rightarrow C_1C_2:V_2$  or lengthening of the following vowel, i.e.  $C_1V_1C_2V_2 \rightarrow C_1C_2V_2:$ . Some examples follow.

(10) Supyire gemination/lengthening after V-loss [N.B: /R/ is a uvular tap (Carlson 1990: 21)]

	<u>C-gemination</u>	<u>V-lengthening</u>	<u>Gloss</u>
pilaga	pl:aRa	pla:Ra	'night'
biliwe	bl:iwe	bli:we	'slave'
kile	kl:e	kle:	'sky, God'
kinaga	kn:aRa	kna:Ra	'fruit bar'
file	fl:e	fle:	'approach'
silege	sl:eRe	sle:Re	'be ashamed'

Hajek and Goedemans (in prep.) argue that the occurring lengthening, either consonantal or vocalic one, is due to the preservation of the mora left behind by the high vowel. They moreover interpret the consonant gemination as a moraic onset geminate, based on two statements of Carlson. First, Supyire has no codas: "There are no closed syllables in Supyire. Syllables are basically CV or CVV. A few grammatical words... begin with a V syllable" (Carlson 1990: 16). As a result, the first part of  $C_2$  cannot be considered a coda. Moreover, Supyire has a single instance of syllabic nasals which arises after the deletion of an unstressed vowel in the environment  $NVC_{[-voi \text{ stop}]}$  (Carlson 1990: 65-66). If these statements are right, then the consonant in question cannot be syllabic or a coda-onset geminate. The only alternative then seems to be that of the consonant being syllabified in the onset and bearing a mora, as a result of mora preservation.

To show that mora preservation is indeed involved, evidence is presented relating to the 1SG gender class  $-wV$  which only attaches to bimoraic roots, such as CVCV *sara-wa* 'bee' or CVV *tfe:-we* 'woman', but not to monomoraic ones. The fact that the suffix can appear after forms generated through elision, like *bli:-we* ~ *bli:we* 'slave' corroborates the idea of mora preservation, since in each case, the resulting root must be bimoraic.

<sup>11</sup> But there are words which conform to this description and yet show no elision, see e.g. *sululi* or *pinini* (Carlson 1990: 231). Perhaps, the vowel in question is stressed and that is why elision does not apply.

But there are certain other details about Supyire's phonology which can make us sceptical about the validity of the above proposal. While mora preservation indeed seems to be occurring, its characterisation may not be accurate. First note that this process only occurs when  $C_2$  is a sonorant /l/ or /n/, and not with any other consonants. Second, the sonorant retains the tone of the elided vowel. Third, given the examples Carlson provides, this process only seems to occur when the relevant sequence appears word-initially.

These three properties are consistent with either syllabic consonants or sesqui-syllables. The latter occur mainly in the Mon-Khmer languages, like Kammu, Temiar or Semai (Svantesson 1983, Sloan 1988, Gafos 1998, Lin 1998, Hendricks 2001 and references cited therein), which possess a large number of words that include a sequence of consonants (usually two), dubbed a minor syllable (or semi-syllable) and a normal syllable, e.g. CC.CVC as in Temiar *br.caa?* 'to feed' (Gafos 1998: 238) or Semai *dh.dnoh* 'appearance of nodding constantly' (Sloan 1988: 320). The first CC part of these words is not really a syllable as it includes no nucleic part, but instead is usually considered as an onset-coda sequence with one mora (hence the name sesqui-syllabic, i.e. words of one-and-a-half syllables; for alternative representations of minor syllables see Lin 1998 and §8.3.2). Such degenerate semi-syllables are confined to word-initial position (Sloan 1988), which according to John Hajek (p.c.) is reasonable given that these languages are very markedly stress-final and all other syllables tend to be reduced. A final property of minor syllables relevant for our purposes is that commonly the second consonant carries tone, e.g. Kammu *k̀h.ní?* 'behind', *h́r.màal* 'soul' (Svantesson 1983: 32-33). Adding all these traits together make minor syllables look suspiciously similar to the  $CC_{[son]}$  sequences of Supyire.

Consequently, we conclude that it is possible to view the geminated consonant being syllabified as  $C_1C_2.C_2V_2$  so that  $C_2$  is either syllabic or a coda of a semi-syllable. The only reason we could not yet adopt any of these two ideas with any certainty is Carlson's statements mentioned above. If the language lacks codas, semi-syllables should not be possible, at least not under the interpretation that views semi-syllables as onset-coda sequences. The other statement maintains that the language has syllabic sonorants, but only in a very specific environment which does not correspond to the one in (10). Here too then syllabic sonorants would not be an option either.

To construct a more informed opinion about Supyire we would then need answers to the following questions: i) on what empirical grounds does Carlson establish that the only syllabic sonorants in the language are the nasals which arise after the

deletion of an unstressed vowel in the environment  $NVC_{[-\text{voi stop}]}$ ? In other words, how can we exclude with certainty that other sonorants cannot be syllabic (or part of a semisyllable)? ii) are there absolutely no examples of the type:  $\text{kapilaga} \rightarrow \text{kap:laga}$  [i.e. with V-elision word-medially]? Hajek and Goedemans seem to suggest that there are none, since onset gemination is today a historic process in the urban variety of Supyire; iii) are there any examples where  $C_2 \neq l, n$ ? And iv) are there any other instances of geminates in Supyire? Before we receive satisfactory answers to these issues, I contend that no decision can be made for one or the other approach.

Prior to closing this section, it is however worthwhile mentioning that there is a small set of data which could be taken to favour the case for onset geminates (Carlson 1990: 59). The description relating to the data in (11) is the following: “Roots ending in stressed CVr or CVN have their vowel lengthened when the diminutive suffix -rV is added. Since only a single consonant remains of the expected cluster  $[rr] \rightarrow [r]$  and  $[nr] \rightarrow [n]$ , this lengthening is probably compensatory following degemination”.

- (11)    cer-    +    -rV    →    ceere  
           calabash    DIM    ‘little calabash’
- ŋkeN- +    -rV    →    ŋkeene  
           branch    DIM    ‘little branch’

The point here is that if there are no codas in the language, then at the stage just before degemination, those geminates should have been wholly syllabified in the onset. Assuming a compensatory lengthening analysis that treats CL as mora preservation, then given that subsequent degemination caused lengthening of the previous vowel, the geminates should carry moras. If this is on the right track, then after all maybe Supyire has - or used to have - moraic onset geminates.

#### 7.2.5.2 Yoruba (Ola 1995, 1997)

This section begins with a disclaimer. Unlike the previous cases, it is quite clear that Standard Yoruba does *not* involve moraic onsets. Nonetheless, reference to it is useful for illustration purposes, as it presents a host of phenomena and processes which in another language could be interpreted as involving moraic onsets. These include: i) Word Minimality where CV words are fine, whereas V ones are not, ii) Intransitive Imperatives of /CV/ verbs which surface as CV without lengthening to CVV or

augmentation to CVCV, e.g. *ba* 'to hide' → *ba* 'hide!' \**baa*; *lɔ* 'to go' → *lɔ* 'go!' \**lɔɔ*,  
 iii) Loan verb truncation, where consonant-initial English loan verbs truncate to the initial CV, e.g. *páàsǐ* → *pá* 'to pass', *póm̀b̀ù* → *pó* 'to pump', and iv) Ideophonic Reduplication which indicates light intensity of action or shape. In this instance the reduplicant surfaces as a single CV suffix, e.g. *rògòdò* → *rògòdò-d̀ò* 'round and big → intensifying'.

One can thus see numerous occasions involving prosodic phenomena where CV syllables are licit. This on its own is not particularly interesting. More revealing is the fact that in the same environments V syllables are not allowed. Thus there are no V words in the language (other than functional words that is); V-initial loan verbs cannot truncate to a single V, but instead present an epenthetic onset *h*, e.g. *énfì* → *hé* but \**é*, \**én* and no V-reduplication arises, e.g. *gbàyàù* → \**gbàyàù-ù* 'open and loose'. The point of interest then is that V syllables behave differently from CV ones. Such an observation seems to be necessary if a case for moraic onsets is put forward (cf. Bella Coola WdMin where CV words but not the V ones are allowed in the language).

However, Ola (1995, 1997) provides robust evidence that a moraic onset analysis is not appropriate, because the generalisation regarding the distinction between CV and V syllables lies elsewhere. The prominent observation is that every Yoruba word needs to contain at least one onset consonant. Thus words that consist of a single vowel whether short or long are banned, i.e. \*V, \*VV. On the contrary, words of the type CV or CVV are permitted. Ola proposes to analyse this pattern by making use of the Properheadedness idea after Itô and Mester (1992). Properheadedness requires that every word must contain at least one foot, every foot at least one syllable and every syllable at least one mora. She then suggests that in Yoruba only CV(V) strings actually make a syllable. Onsetless syllables are not possible, therefore V or VV are unsyllabified<sup>12</sup>. Nonetheless, such sequences are licensed by moras (in a manner similar to Bella Coola moraic licensing) and hence can surface.

Properheadedness is argued to be a top priority in the language, which accounts for the facts above, where CV syllables but not V ones are allowed. A striking manifestation of this insight comes through the consideration of consonant deletion processes (Ola 1995: 277-8 and references therein). There are three basic types of consonant deletion, namely (i) sonorant deletion, (ii) deletion by identity, (iii) /r/ deletion. We will only consider the first. The other two are similar.

<sup>12</sup> This is actually a generalization pertaining to Standard Yoruba. Ondo Yoruba for instance (Ola 1997) possesses no similar restrictions, since V and CV syllables are treated the same.

Sonorant /r/ and glides optionally delete when they are intervocalic. This deletion is iterative in the sense that if there are two sonorants in the word, both may delete. There is just one case where deletion is blocked and that is when it would result in a word that had no consonant onset at all, e.g. as in \*eúé below. Without blocking of deletion, the word would effectively be rendered syllable-less and thus violate the Properheadedness requirement. Some examples are given below.

(12) *Optional Intervocalic Sonorant Deletion*<sup>13</sup>

- a. ewúré ~ euré ~ ewúé ~ \*eúé      'goat'
- b. àwúré ~ àuré ~ àwúé ~ \*àúé      'luck charm'
- c. ìrágberí ~ àágberí ~ àágbeé      'city name'

Despite showing that Yoruba does not fall within the class of languages that present onset moraicity effects, we can nonetheless - given the language's facts - easily imagine a language similar to Yoruba that would belong to the moraic onset category. Such language would be like Yoruba in that it would distinguish CV from V sequences for a host of phenomena, but it would either lack processes like the deletion pattern above or - even preferably - exhibit processes of this type too, but without the 'at least one onset per word' restriction Yoruba has<sup>14</sup>. It remains to see whether there is actually a language that fits the bill.

### 7.2.5.3 *Tukang Besi (Donohue 1999)*

Tukang Besi, a language of Southeast Sulawesi, Indonesia illustrates a case where *alleged* geminates syllabify wholly in an onset position, but provides no clear support for onset moraicity. Probably these are then more accurately described as 'doubled consonants with two root nodes' (§6.5 and especially §6.5.2). Some discussion will clarify why this is the case.

<sup>13</sup> One might be tempted to claim that the reason \*eúé or \*àúé are bad is because they involve super-long vocalic hiatus, which is generally avoided in languages (Kavitskaya 2002 and chapter 5 here). This is not a particularly promising explanation for two reasons; first there are Standard Yoruba words with exactly this structure, e.g. in the reduplicated form òrò-òórò 'every morning'. Still the fact that there is a morphological boundary between the vowels may play some role. The potential dislike for super-long vocalic hiatus cannot account for the following facts though which arise in another type of consonant deletion, i.e. that of /r/-deletion. As before, /r/-deletes whenever there is still another onset in the word, e.g. erùpè ~ èpè 'sand', but not when this would lead to an onsetless word, e.g. orí ~ \*of 'head', àrà ~ \*àá 'thunder'. Here the resulting vowels are simply long - which are very common in Yoruba - and yet certain deletions are blocked. The requirement against words that totally lack syllables captures this pattern and is consistent with what happens in (12).

<sup>14</sup> In that case, we would also expect that unlike Yoruba, VV minimal words should be accepted too, as the requirement imposed would bear on moraicity and not on properheadedness.



The phonemic inventory includes voiceless and voiced obstruents (although most voiced stops only emerge in recent borrowings), glottals, implosives, nasals, liquids and prenasalized stops (voiced and voiceless ones). Syllables in this language are simply (C)V. This statement is consistent with the fact that the language allows no codas and bans consonant clusters.

However some may view prenasalized stops as clusters occupying a coda-onset sequence. Donohue (1999) argues that this cannot be the case. Instead, he shows that prenasalized stops phonologically act as a single segment and occupy the onset of a syllable. First, native speakers unanimously syllabify a prenasalized stop in the onset of a syllable. Even if these complex segments were clusters, then it is quite peculiar why other clusters such as ordinary complex onsets of the type *tr*, *kl*, fail to arise. The most convincing evidence however comes from reduplication, where the reduplicant consists of two syllables as in (13a).

(13) *2σ reduplication in Tukang Besi*

a. hesowui	→	heso-hesowui	'wash → wash playfully'
b. karambau	→	kara-karambau	*karam-karambau 'buffalo'

(13b) is the more interesting case, which includes a prenasalized stop. If this were a cluster with a coda-onset syllabification, then we would expect a reduplicated form *\*karam-karambau*. The fact that we only get *\*kara-karambau* suggests that the sequence is syllabified wholly in the onset<sup>15</sup>. Bantu languages with prenasalization lead to the same conclusion (cf. for instance the discussion on Kinyarwanda, §5.3).

All this can be tied in in the case of alleged gemination. Tukang Besi optionally seems to geminate consonants in the stressed position of the word, which happens to be the penult (14a). In a disyllabic word, where the stressed syllable is also the initial one, gemination jumps one syllable forward (14b).

(14) *Optional gemination in the stressed syllable*

a)	top:ɔŋa	'cut branches'
	mel:ai	'far'

<sup>15</sup> The die-hard supporter of a coda-onset analysis could claim that in OT this would merely present a TETU effect (Emergence of the Unmarked, McCarthy and Prince 1994), where the language generally admits codas, but not in the reduplicants where the unmarked situation arises, i.e. no codas. There is a way to test the TETU effect: if homorganic clusters are allowed, then for the base /tumpe/ 'first born' the reduplicant could be either: [tumpe-tumpe] (no TETU) or [tupe-tumpe] (TETU). Mark Donohue (p.c., 21/03/06) verifies that the first actually occurs as we get [tumpe-tumpe] for /tumpe/, [pindi-pindi] for /pindi/ 'firm excrement', but that also initial NC stops are preserved, e.g. [ndanga-ndanga] 'jackfruit'.

b)	kap:i	'wing'
	ek:a	'climb'

However if it is true that *Tukang Besi* lacks codas, then these consonants have to be syllabified wholly in the onset position, but since we cannot say much with respect to their moraicity, it is at this stage safer to assume these are doubled consonants rather than geminates.

### 7.3 Misanalysed onset sensitive languages

In this section, I focus on languages, which according to some, present onset weight effects with respect to geminates or stress. I attempt to show that these analyses take no notice of certain data and facts in the languages at hand. As a result, wrong conclusions are reached which involve claims about onset moraicity. In most cases, the languages are well-behaved (in terms of stress: *Bislama* and *Nankina*) or slightly unusual (i.e. *Berawan* *onset* geminates), but at any rate provide no evidence for onset moraicity. I first consider gemination in *Berawan* and then look into *Bislama* and *Nankina* stress.

#### 7.3.1 *Berawan* (García-Bellido and Clayre 1997, henceforth GB&C)

*Berawan* is an Austronesian language spoken in the island of Borneo, in Sarawak, East Malaysia. It is spoken by around 3,000-4,000 people. The dialect described in GB&C is the one of Long Terawan on the Tutoh river. The data reported have been collected by Beatrice Clayre and reflect the speech of Denny Belawing Wan, a native of Long Terawan.

The aspect of *Berawan* which is of interest to us is the existence of geminate consonants. Their presence has been confirmed by acoustic studies that the authors have conducted as well as by investigations by other researchers too, i.e. Robert Blust and Iovanna Condax.

All consonants - plosives, affricates, nasals and liquids<sup>16</sup> - apart from the fricative *s*, the approximants *w* and *j* and the glottals *ʔ* and *h*, can geminate. Geminates have a very restricted distribution. Independently of the analysis proposed, (at least) the second half of the geminate can only emerge as the onset of the final syllable in a word.

---

<sup>16</sup> The authors represent the singleton vs. geminate rhotic as /r/ and /r/ respectively.

(15) *Geminates vs. singletons in Berawan* [GB&C: 25]

<u>Geminate</u>		<u>Singleton</u>	
lut:ɔʔ	‘float’	lutɔh	‘soggy’
ʔɛtʃ:ɔh	‘dog’	ʔɛtʃo:	‘remove’
sɛn:ɐiʔ	‘heat of sun’	sɛnɐiʔ	‘insect sp.’

More specifically, in GB&C's description, the "heavy consonants can only occur as onsets of the nuclear syllable (p. 25)", i.e. the final syllable. This syllable is called nuclear as it is the only one which can support the full inventory of vowels (apart from extra light schwa) and consonants (long and short). Moreover, it is the one that receives the nuclear high pitch H\* of the rising pitch LH\* characteristic of Berawan words. Less explicitly, GB&C (1997: 27) seem to argue that stress too appears on the final syllable of the word.

Typically, words consist of the nuclear syllable and are optionally preceded by one or two light syllables. In fact, the major argument for both the heaviness as well as the syllabification of geminates wholly in the final syllable comes from the observation that disyllabic words are iambic (◡—), while the maximum weight expansion of the word is an anapaest (◡◡—). Had the geminates been heterosyllabic, then the penultimate would be rendered heavy and thus the above generalisation would be lost. The fact that penultimate syllables never comprise a VV or VC either, i.e. they are never heavy, also fits nicely with this pattern. The proposed representation then is one where the geminate is tautosyllabic and gets syllabified wholly in the onset of the last syllable as in (16b) [N.B: I use here C<sub>1</sub>C<sub>1</sub> and C<sub>2</sub> interchangeably to refer to geminate consonants].

- (16)    **a.** Heterosyllabic geminate syllabification      **b.** Tautosyllabic syllabification  
VC<sub>1</sub>.C<sub>1</sub>V                                  V.C<sub>1</sub>C<sub>1</sub>V

It is thus no coincidence that only the final syllable can host a geminate. It is the only syllable that supports a heavy syllable. Had the other syllables included a geminate, then they would become heavy, a fact banned in the language.

Obviously, this argumentation sounds a bit circular, while it heavily draws on a rather implausible assumption, namely that the maximal weight pattern in the language is an anapaest. Anapaests have no theoretical import although they are common in poetic meter, so it is not obvious why it should be profitable to construct an account based on them. For the sake of the argument however, I will attempt a sketchy OT analysis of these data based on the ideas of GB&C.

Given that LH iambs or LLH anapaests are preferred, I will build an analysis where the main idea is that each word has only one heavy syllable and this is aligned with the right edge of the word.

- (17) ALIGN-FtHD-R: Align the right edge of the foot head with the right edge of the word

S(tress)-T(o)-W(eight) can be employed to force stressed syllables to be heavy.

- (18) STW: Stressed syllables are heavy

Combining these two will produce the right effects and will also cause geminates to syllabify as tautosyllabic in the final syllable. As a result, no reference to anapaests is required. Suppose we need to parse an input sequence such as /L L H/. The winner is indeed an LH iamb, as it manages to satisfy both constraints at the same time, unlike its rivals.

- (19) ALIGN-FtHD-R, STW [stressed syllables are underlined]

/L L H/	ALIGN-FtHD-R	STW
a. L (L <u>H</u> )		
b. L ( <u>L</u> H)	*!	*
c. L ( <u>H</u> ) H	*!	
d. L ( <u>H</u> ) ( <u>H</u> )	*!	

This ranking does not suffice once we consider an input like /LHH/ (20). The problem is that both candidates here fare equally well. However, adding the WSP and ranking it over MAX-IO- $\mu$  settles this glitch (21), by getting rid of (b), whose penult remains unstressed albeit heavy and favouring (a) which satisfies WSP by shortening the penult.

- (20) ALIGN-FtHD-R, STW

/L H H/	ALIGN-FtHD-R	STW
a. L (L <u>H</u> )		
b. L H ( <u>H</u> )		

- (21) ALIGN-FtHD-R, WSP, STW >> MAX-IO- $\mu$

/L H H/	ALIGN-FtHD-R	WSP	STW	MAX-IO- $\mu$
a. L (L <u>H</u> )				*
b. L H ( <u>H</u> )		*!		

Importantly, assuming that STW also dominates DEP-IO- $\mu$ , then the ranking will also force a /LLL/ input to emerge as [LLH].

(22) ALIGN-FtHD-R, WSP, STW >> DEP-IO- $\mu$

	/L L L/	ALIGN-FtHD-R	WSP	STW	DEP-IO- $\mu$
a.	L (L <u>H</u> )				*
b.	L (L <u>L</u> )			*!	

We can now examine what happens when the input includes a geminate consonant. The question is how this is going to syllabify. All we need to do is to add the generic \*MORAIC ONSET which would penalise tautosyllabic geminates and rank it below the other high-ranking constraints used before. As shown in (23), (a) loses because by syllabifying the geminate heterosyllabically, it assigns the mora on the penult leaving the final stressed syllable light. This causes a fatal violation of STW. (b) has the same syllabification but stresses the penult, which although satisfies STW, drags stress away from the right edge of the word and produces a violation of dominant ALIGN-FtHD-R. (c) satisfies both high-ranking constraints by syllabifying the geminate in a more marked configuration, namely that of a tautosyllabic geminate<sup>17</sup>.

(23) ALIGN-FtHD-R, STW >> \*MORAIC ONSET

	/CV $\mu$ C $\mu$ V $\mu$ /	ALIGN-FtHD-R	STW	*MORAIC ONSET
a.	CV $\mu$ C $\mu$ <u>C<math>\mu</math></u> V $\mu$		*!	
b.	<u>CV<math>\mu</math></u> C $\mu$ <u>C<math>\mu</math></u> V $\mu$	*!		
c.	CV $\mu$ . <u>C<math>\mu</math></u> V $\mu$			*

While this analysis would require modifications to account for the full range of facts, the point here is to show that an analysis along the lines of GB&C is possible. However, there are reasons to question the validity of this account. Some are presented immediately below.

Berawan has a Word Minimality condition which takes the form CVV or CVC<sup>18</sup> (p. 27-28). This is obviously a bimoraic requirement, which naturally raises the following question. If geminates are heavy, then why doesn't a C:V fulfil the bimoraicity requirement? There is a possible way out. Given that geminates only occur

<sup>17</sup> We would like to be able to generate an output like [(CVCV:)], even if the input had been something like /C $\mu$ VCV/. To achieve this, we would also need ALL-Ft-R to dominate all other constraints so that the appealing candidate [(C:V)(CV:)] is ruled out. The rest of the ranking would favour the output [(CV.CV:)]

<sup>18</sup> A minimal syllable contains an onset (GB&C: 36).

in the ultima of polysyllables, we could argue that initial geminates are banned. As a result C:V words would be bad, simply because they would include the prohibited initial geminates. But no similar rescue is available once we consider the full inventory of word-final syllables.

(24) *Distribution of word-final syllables*<sup>19</sup>

N.B: brackets refer to the original numbering of examples in GB&C and not to page numbers

a. open syllables

\*V

\*CV

CV:	bi:	“lips” (GB&C: (1))
CV <sub>1</sub> V <sub>2</sub>	saʔaɪ	“frog sp.” (GB&C: (1))
CV <sub>1</sub> :V <sub>2</sub>	sapa:u	“roof” (GB&C: (1))
CV <sub>1</sub> V <sub>2</sub> :	kio:	“follow” (GB&C: (4))

b. closed syllables

\*VC

CVC	pukəʔ	“jungle knife” (GB&C: (7))
CV:C	tuto:m	“joint” (GB&C: (35))
CV <sub>1</sub> V <sub>2</sub> C	buaŋ	“beetle” (GB&C: (4))
*CV <sub>1</sub> :V <sub>2</sub> C		
*CV <sub>1</sub> V <sub>2</sub> :C		

c. syllables with geminates

\*C:V

C:V:	dʒik:ɑ:	“rendezvous” (GB&C: (1))
C:V <sub>1</sub> V <sub>2</sub>	ʔap:u	“height” (GB&C: (25))
C:VC	bəl:ɪn	“hill” (GB&C: (9))
C:V:C	ŋəp:ɑ:ŋ	“wasp sp.” (GB&C: (9))
C:V <sub>1</sub> V <sub>2</sub> C	sap:auʔ	“blowpipe” (GB&C: (4))

d. exception (possible loanword):

CV <sub>1</sub> V <sub>2</sub> V <sub>3</sub>	bəliau	“shaman” (GB&C: (4); see GB&C fn. 4)
---	--------	--------------------------------------

These data are very informative. First, we can see that onsets are obligatory in monosyllabic words ruling out forms such as [VV] and [VC]. While this excludes [V] words too, their absence undoubtedly relates to the bimoraicity condition. Consequently, [CV] words are bad even though they are onsetful. As we have seen already, [C:V] words are prohibited, despite being both onsetful and allegedly bimoraic. Recourse to the lack of initial geminates can account for the absence of such words. The

<sup>19</sup> Following García-Bellido and Clayre, I use [ɑ:] for the long low central vowel, but [a] for its short slightly fronter version, rather than the [ɐ] they use.

problem however arises in larger words, where the geminate would no longer be word-initial. In other words, we should anticipate forms such as [CV.C:V], because assuming the geminate is heavy and syllabified in an onset, then this word would be a well-formed [LH] iambic sequence. And yet, GB&C offer no example of this type.

If however we accept that the final syllable must indeed be heavy (although no similar expectations apply to other syllables) and that so-called geminates are in fact weightless doubled consonants (§6.5.2), then facts fall into place. The reason we do not get [CVC:V] is because this would actually be [CV<sub>μ</sub>.C:V<sub>μ</sub>], i.e. \*[LL], which is forbidden since the last syllable needs to be heavy. At the same time, configurations such as CV.C:V: e.g. dʒik:ɑ: “rendezvous” (GB&C: ex. 1, p. 18) or CV.C:VC e.g. lak:ɛh “omen bird” (GB&C: ex. 9, p. 25) would be the only possible, because they could render the final syllable heavy due to the bi-moraicity of the rime.

A further problem with pursuing the idea that there are moraic onset geminates in Berawan is the presence of C:V:C syllables as in e.g. [ŋəŋ:ɑ:ŋ]. If geminates contributed a mora, then this syllable which includes a bimoraic nucleus and a moraic coda should be *quadrिमoraic*. This is totally unprecedented. In fact, GB&C note that the maximum of the moras in the *rime* is three (p. 36) and make no reference to the prediction about quadrिमoraic weight that is unavoidable in their analysis. Positing three moras as the maximum *syllable* weight however is consistent with the presence of C:V:C syllables (where the first consonant is an initial doubled C) as well as with the lack of \*CV<sub>1</sub>:V<sub>2</sub>C and \*CV<sub>1</sub>V<sub>2</sub>:C that would exceed this maximum.

All these problems indicate that so-called Berawan geminates are not in fact geminates. There is no evidence for their moraicity, therefore they should be considered ‘doubled consonants with two root nodes’. This proposal also explains why \*C:V syllables do not satisfy Word Minimality, why final syllables which are always heavy never include a C:V syllable (as this is light) and why seemingly quadrिमoraic syllables are allowed. The latter are permitted simply because they are actually trimoraic. Trimoraic syllables are marked, but extant in a few languages, e.g. Estonian (Hayes 1995) or Kashmiri (Morén 2000).

At this point, one question comes up. How do we account for the absence of doubled consonants in other syllables apart from the final one, where contrasts such as the following can be found e.g. lut:ɔʔ “float” vs. lutɔh “soggy” (GB&C: (25))? Obviously this relates to a better understanding of the properties of doubled consonants, which at present remains vague. While I will currently set this aside, it is perhaps

instructive that the final syllable in Berawan is the nuclear syllable, the one that bears nuclear pitch and stress. We could perhaps then argue that these consonants are somehow licensed by stress, shaping up an implicational relationship, namely: ‘if there is a doubled consonant, then it is in a stressed syllable’, while the reverse does not hold.

As a final remark, observe that Berawan resembles Trique (§7.2.2.3) in that final syllables license certain contrasts and allow lengthened consonants to appear. However, in Trique I reached the conclusion that moraic onsets were involved, which has not been the case for Berawan. The reason for that is that an analysis that employs moraic onset geminates in Berawan, leaves - as I have just demonstrated - numerous aspects of the language unaccounted for, i.e. WdMin considerations, final syllable patterns and weight restrictions. On the other hand, the available Trique data seem supportive of and compatible with the use of moraic onsets.

### **7.3.2 Bislama (Camden 1977, Gordon 2005, Lynch 1975, Crowley 2004)**

Bislama is the national language of Vanuatu, which started as a plantation pidgin in the nineteenth century, but since then has evolved into a fully-fledged language. It has significantly diverged from English and French upon which large part of the vocabulary is based. Bislama is relevant here since Gordon (2005) citing Camden (1977) argues that it presents a stress algorithm which is based on the following weight scale:

(25) *Bislama weight scale (Gordon 2005)*

CCVC > CVC > CCV > CV

I will show that this claim is not supported by any evidence. While not much work has been done on Bislama stress, it most likely seems to be a system with partially unpredictable stress, although the preferable pattern involves penultimate stress.

First, let us consider the Bislama stress description in Camden (1977: xiv-xv). An obvious problem is that this is presented in merely two pages, but most crucially no examples are given. This generates an important question. Is the stress algorithm presented in Camden an accurate description of the stress system of Bislama or did Camden overgeneralise claiming broad patterns on the basis of his limited data? In Appendix B, I show that most of the time one can find examples that would correspond to the patterns Camden presents, but of course there is no indication of stress which



would allow us to form any generalisations. Moreover, no discussion is reserved for possible effects of morphology. In short, Camden's description goes like this:

- (26) two-syllable words: penult stress
- exception: reduplication with output 'CVC' CVC
- (27) three-syllable words: penult stress. But final stress if:
- i) final  $\sigma$  = CVC
    - except in:  $\sigma \cdot (C)(C)(C)VC.CVC$  which includes more specific words of the type:
      - a) CCV.'CCVC.CVC
      - b) CCCV.'CCCVC.CVC
      - c) CCCV.'CCCV.CVC
      - d)  $\sigma \cdot$  CVC.CVC
  - ii) final  $\sigma$  = CCV
    - except in:
      - a) V.'CVC.CCV
  - iii) final  $\sigma$  = CCVC

Other exceptions in trisyllabic words involve:

- (28) V.CV.'VC
- (29) CVC.V.'VC [or CV.CV.'VC]<sup>20</sup>
- (30) CCVC.V.'CV [or CCV.CV.'CV]
- (31) four syllable words:  $\sigma\sigma\sigma\sigma$ , but final stress if:  
 $\sigma\sigma\sigma \cdot$  CVC, except penult and ultima are: 'CVC.CVC
- (32) five syllable words:  $\sigma\sigma\sigma\sigma\sigma$ , but final stress if:  
 $\sigma\sigma\sigma\sigma \cdot$  CVC
- (33) six syllable words:  $\sigma\sigma\sigma\sigma\sigma\sigma$

Gordon (2005) claims that only the two final syllables can attract stress and based on Camden's description (27ii) he infers that CCV syllables seem to attract stress, but not

<sup>20</sup> The syllabification in brackets is the one we would expect given standard assumptions about syllabification.

as much as CVC ones, since by (27ii.a)  $V.\acute{CVC.CCV}$ , we can see that  $CVC > CCV$ . Statements (28) and (29) also show that  $VC > CV$ . Adding all this together, Gordon concludes that  $CCVC > CVC > CCV > CV$ . Thus, stress surfaces either on the penult or final with the penult being the default. Stress is on the final when the ultima is heavier than the penult. Closed syllables are heavier than open ones and complex onsets are heavier than simplex onsets. However, the rime weight has precedence over onset weight, which explains why  $CVC > CCV$ .

As we have seen however, Camden's description raises issues of reliability as no examples are given. Moreover, his syllabification seems at least peculiar. Consider for instance statements (29)-(30), all of which include the initial sequence (C)CVCV. Camden syllabifies this as (C)CVC.V rather than the more usual (C)CV.CV. He provides no justification why this should be the case, but this can make us suspicious of the validity of the syllabification overall.

Setting this matter aside and assuming for the sake of the argument that the approaches of both Camden and Gordon are on the right track, then their predictions diverge in many respects. First, Camden says that final CCV takes stress, unless  $V.\acute{CVC.CCV}$  (27ii.a). If we interpret Camden as saying that all words [in the language not in his data only] with final CCV get stress apart from this one exception then this distinguishes him from Gordon who would expect final stress only when the penult is V or CV, as shown below:

- (34) Camden vs. Gordon with respect to CCV final syllables [in bold the cases where the predictions differ]

<i>Camden</i>	<i>Gordon</i>
$\sigma.V.\acute{CCV}$	$\sigma.V.\acute{CCV}$ [ $CCV > V$ ]
$\sigma.CV.\acute{CCV}$	$\sigma.CV.\acute{CCV}$ [ $CCV > CV$ ]
$\sigma.\acute{CVC.CCV}$ Exception	$\sigma.\acute{CVC.CCV}$ [ $CVC > CCV$ ]
$\sigma.CCV.\acute{CCV}$	$\sigma.\acute{CCV.CCV}$ [default]
$\sigma.CCVC.\acute{CCV}$	$\sigma.\acute{CCVC.CCV}$ [ $CCVC > CCV$ ]
$\sigma.CCCV.\acute{CCV}$	$\sigma.\acute{CCCV.CCV}$ [ $CCCV > CCV?$ ]
$\sigma.CCCVC.\acute{CCV}$	$\sigma.\acute{CCCVC.CCV}$ [ $CCCVC > CCV$ ]

In addition, in  $4\sigma$ - $5\sigma$  words, stress is normally on the penult, but Camden has a special treatment for CVC final words which receive stress on the ultima (cf. (31) and (32)). But if according to Gordon, other types such as CCVC and CCV are heavy, then we should get similar results, i.e. word-final stress, whenever CCVC or CCV are final and the heaviest in the disyllabic window.

Moreover disyllabic words according to Camden receive penultimate stress. If the hierarchy that Gordon proposes holds, we would expect words with heavy ultimas to be stressed word-finally rather than in penultimate position. Many examples where this could be seen are those where the ultima is CCV or (C)(C)(C)VC [and the first syllable is lighter]. Camden states that the first syllable gets stress, Gordon predicts that the second does.

While it remains to see whether any of these predictions is confirmed, we should also draw our attention to other descriptions that make both of these accounts unlikely or at the very least questionable. Lynch (1975: 193-5) observes that almost all disyllables, quadrisyllables and the majority of trisyllables stress the penult. Quadrisyllables also have a stress on the peninitial. Some exceptions occur in words which include morphemes like -ap (stress on final), e.g. *antáp* 'up, above', *klosáp* 'close, near' or -fala ~ -vala (stress on initial), e.g. *mívala* 'we (exclusive)', *bikfala* 'big (emphatic)'. Still others have retained the stress of the source language, e.g. *kámbani* 'group', *évriwán* 'everyone' (from English) or present initial stress, albeit being native ones, e.g. *nákamal* 'dance-ground, kava-drinking place, men's house'<sup>21</sup>.

This description is more in accordance to a recent discussion of Bislama stress by the late Terry Crowley (Crowley 2004). Stress is claimed to be unpredictable, although some strong tendencies can be identified. Words of Melanesian origin overwhelmingly present stress on the penult, as it happens to be the case in the source languages. Words of French descent often have stress on the final syllable, whereas those of English origin usually keep the stress they present in English. Of course, Crowley emphasises that these are strong tendencies, but exceptions occur, which justify the partial unpredictability of the Bislama stress system. It is interesting however to mention that in personal communication with Crowley (22/03/04), he notes that some representative words like *épril*, *fútbrek*, *dívos*, *baot*, all present penult stress. Obviously *épril* contradicts both Camden and Gordon, since the final syllable is by both accounts a heavy one and yet receives no stress. Similarly for *fútbrek*, *dívos*, *báot*. While these are loanwords and their stress could be attributed to effects from the source language, such issues do not arise in either Camden or Gordon who propose algorithms that take no account of the loanword phonology.

One thing is for sure; while there is no consensus on the Bislama data, recent accounts of the facts in Crowley (2004) do not suggest any effect from the onsets. Until

---

<sup>21</sup> Note that according to both Gordon and Camden, a word like that should present final stress.

more robust evidence arises in favour of onset effects on Bislama stress, I believe it is safer to assume that this language has a quite common system with penult stress and partial lexical stress as a result largely of the heavy borrowing from English, French and other Melanesian languages.

### 7.3.3 Nankina (Spaulding & Spaulding 1994, Al-Harbi 2005, Gordon 2005)

Nankina is a language of Papua New Guinea spoken in the region of the Madang Province, Saidor District, in the upper Nankina River valley. According to Gordon (2005) the language's stress system provides evidence for CCV > (C)V, that is, syllables with complex onsets act as heavier than ones which have a simplex onset or lack one altogether, and thus CCV syllables attract stress more. Coda consonants do not count for weight purposes. While Al-Harbi (2005) correctly argues that CCV syllables do not behave as heavier than (C)V ones, I contend that his conclusion is right but for the wrong reasons, so that the proposed analysis does not really show that onsets cannot contribute to syllable weight, simply because Nankina does not lend itself as testing ground for such a hypothesis. This is because Al-Harbi starts from the assumption that Gordon is right in the interpretation of the data, but as I will show, this is misguided. Let us first examine the data more closely<sup>22</sup>.

Nankina stems usually consist of one or two syllables, so stress normally docks on the first syllable.

#### (35) Normal - penultimate stress

tó.wuŋ	'egg'	wó.re	'a sore'
ká."dep	'wood'	jé.ɣi	'cause'

Syllables with [ɪ] or [i] nuclei are considered extra light, thus when they occur word-initially, stress is found on the second syllable.

#### (36) Extra light syllables - final stress

<sup>m</sup> bi.ts <sup>j</sup> ép	'time'	jɪp.mán	'let go'
------------------------------------	--------	---------	----------

The contentious data are the following:

<sup>22</sup> Spaulding and Spaulding mention a few other minor patterns where for instance syllables receive equal stress when their nuclei are identical, e.g. tsá.wát 'machette', áp.mák 'cough' or stress might be slightly complicated depending on whether there is reduplication. For more details, see (1994: 18-19).

(37) CCV attracts stress - final stress

ε.kwí                      'bad'                      ε.ts'éŋ                      'light weight'

Al-Harbi considers both the vowels [ɪ] and [i] as well as onsetless vowels as weightless. Moreover, he claims that these are distinguished from one another, but this is never explained. Although the whole analysis seems precarious, the gist of the idea is that some notion of foot-head *prominence* dictates that syllables with the above characteristics are avoided for stress placement.

The first problem is the acceptance that Nankina indeed has complex onsets. Spaulding and Spaulding (1994) from the very beginning state that word-initial clusters such as /pt/, /bt/ or /kt/ are split by what they call 'a transitional vowel [i]' whose exact phonetic realization varies. Thus one finds: /pta/ - [pɪra] 'stand', /ktagwək/ - [kɪra<sup>n</sup>gwək] 'small knife', /bt/ - [bɪt] 'pig'. Nothing like that happens in cases like /kwit/ - [kwit]/\*[kiwit] 'bird' or /mwək/ - [mwək]/\*[miwək] 'lizard'.

The same epenthetic vowel also appears in Pidgin words that clearly involve complex onsets:

(38)	<u>Nankina</u>	<u>Pidgin</u>	<u>Gloss</u>
	giras	gras	'grass, hair'
	pires	ples	'village'
	sitori	stori	'story'
	birupera	blupela	'blue'

All these suggest that Nankina lacks true complex onsets. It only presents affricates, prenasalised segments, and consonant-glide sequences (with the most popular being /kw/ and /gw/). All these can be considered complex segments with complex or secondary articulation, but the point is that there is just one root node involved, therefore they are not true complex onsets.

On the other hand, occasional sequences such as /pm/ in ʔp.mʌk 'cough' are correctly treated in Spaulding and Spaulding as coda-onset ones. Unfortunately though, in the phonology part of their grammar they are silent about what happens to sequences that are not obvious loanwords and yet could easily syllabify as complex onsets. For instance one finds *kipra* 'dry' (1994: 78) or *bra* 'hold' (1994: 177) in later morpho-syntactic sections of the grammar where no phonetic transcription is shown. I would guess that these are probably split by [i] too, but empirical confirmation is required.

As we have seen in (37), alleged CCV syllables attract stress, but only when the competitor syllable for stress is onsetless. If CCV is, as we claim, truly CV, then the V.'CCV pattern actually translates to V.'CV, which of course brings us to the familiar pattern of onsetful syllables attracting stress (like in Alyawarra, Aranda and many others, cf. §3.3.1). In fact, Spaulding and Spaulding (1994) also present examples of the V.CV(C) type [(39)], where stress is placed on the final syllable like in (37).

(39) CV syllables attract stress when preceded by onsetless ones

ájét	'louse'	ájún	'meat'
------	---------	------	--------

Al-Harbi chooses to categorize these data as having stress on both the first syllables. While this is indeed the transcription the original data in Spaulding and Spaulding have, it seems to me that this is a slightly different issue. The main generalisation is that onsetful syllables are preferred for stress placement over onsetless ones - which only appear word-initially - whereas stress on the onsetless syllable may relate to different factors which are yet to be established. After all, this 'double-stressing' also occurs in certain cases of reduplication or when the two first syllables have identical vowels (see fn. 22). I thus contend that it would be wrong to ignore a frequent generalisation and prefer to set aside data as the ones in (39).

The most convincing evidence though is the following. The word *takwʌn* is stressed like [tákʷʌn] and not \*[takwʌ́n], as one would expect given Al-Harbi's and Gordon's analyses. Suppose as Gordon and Al-Harbi do that the language indeed has complex onsets. Then for Gordon, final stress would be predicted, while for Al-Harbi, things are a bit more complex, since he does not have anything to say for the relationship between CCV and CV. But if one drives his point to its logical conclusion, then it is likely that CCV syllables are considered more prominent than CV ones, although we cannot be sure since there is no explanation on what 'prominent' exactly means in that analysis. In that case then, Al-Harbi would agree with Gordon, which is of course the wrong prediction.

Alternatively, suppose that a string like /kw/ is really a single segment, albeit a complex one. Then it is only natural that the word should be stressed like [tákʷʌn], since this is really a CV.CV(C) sequence. In other words, the syllables are equally light, so initial stress is chosen, since this is the normal pattern the language chooses, when no other pressing requirement is applicable. Consequently, Nankina does not provide support for CCV > CV. Even if we accept that complex onsets are allowed - despite the

evidence that they are not - there can be no explanation why CV and CCV act in the same way, other than that the language does not distinguish between the two. The language's basic stress algorithm can be simply put as: "Stress the leftmost syllable. Exceptionally, the peninitial syllable receives stress when the initial: (a) includes a stress retracting vowel such as [ɪ] and [i], (b) includes an onsetless syllable". The stress retraction pattern of (a) can easily be accounted for under a theory like Kenstowicz (1994b) and de Lacy (to appear) where head positions avoid nuclei of low sonority such as those of central or high vowels. (b) is also a common pattern where stress is aligned with the first onsetful syllable, as we have seen in Chapter 3.

We have thus offered evidence that both alleged cases of CCV > CV (Gordon 2005) as manifested in Nankina and Bislama cannot be justified. As things stand, no language seems to distinguish between complex and simplex onsets in terms of weight contribution<sup>23</sup>. This is an issue which I will briefly address again from a different perspective in the next and final chapter. In that chapter, I will summarize the main points of the thesis, test some of the claims in a tentative phonetic experiment and draw attention to issues that call for future investigation.

---

<sup>23</sup> Recall that Luganda metrics seems to pay attention to some complex onsets, but this requires further investigation, since far less data are available for this phenomenon.

## Chapter 8

### Concluding remarks and directions for future research

#### 8.1 Concluding remarks

This thesis has explicitly argued that we should take seriously the fact that on the *observational level* onsets are prosodically active. While most theorists in the past bypassed this fact by merely acknowledging it (Hayes 1995, Gordon 1999, Morén 1999/2001) and “conveniently” (Morén 1999/2001: 7-8) setting it aside, I have instead tackled it directly and have claimed that allowing weight in onsets accounts for a host of diverse phenomena ranging from stress to word minimality, compensatory lengthening and gemination among others.

All previous proposals (Everett 1988, Hayes 1995, Goedemans 1998, de Lacy 2000, Smith 2005, Gordon 2005) have only focused on a very specific phenomenon relating to onsets, that of onset-sensitive stress. Using data from Pirahã (Everett and Everett 1984, Everett 1998) and Australian languages like Aranda (Strehlow 1944) and Alyawarra (Yallop 1977), they have built *prominence* accounts in which onsets or certain types of onsets attract stress on the syllables that host them. In §1.3, I have reviewed many of these proposals in detail and have highlighted certain shortcomings each bears.

Most crucially however, the major defect that *all* of them share constitutes the most important argument for an account along the lines proposed in the current work. Prominence accounts are inherently designed to explain phenomena that are unambiguously or under at least some interpretation, prominence-based. Stress is obviously one of these. Consequently, prominence may be able to successfully account for onset-sensitive stress, but this is as far as it can go. It falls short as soon as it encounters phenomena which are unquestionably weight-based, such as word minimality, compensatory lengthening, gemination, reduplication, and others. As I have shown, for all these phenomena, there exist onset-sensitive instantiations, e.g. Bella Coola word minimality (Chapter 4), Samothraki Greek compensatory lengthening (Chapter 5), Pattani Malay, Trukese and Marshallese geminates (Chapter 6) and numerous others (Chapter 7).



An analysis that admits onset weight can account for all these phenomena, but it can also account for onset-sensitive stress, implying that prominence is not the right theory for (the overwhelming majority of) prosodic effects relating to onsets. Nevertheless, in Chapter 1, I have argued that with regard to onset-sensitive stress, a further distinction needs to be made. While most researchers have collapsed the effect the presence of an onset may have with that of the quality of the onset, I have suggested that these are separate phenomena as shown in (1). I have claimed that the ‘quality of onset’ phenomenon pertains to moraic onsets (chapter 2; pattern ❸), whereas the ‘presence of onset’ phenomenon does not involve moraic onsets (chapter 3; pattern ❹), but can be analyzed in terms of alignment (Goedemans 1996) or even prominence as Smith (2005) suggests. Given that these two are independent from one another, they can also interact, as schematized in cell ❶ (Chapter 2). Of course, no onset effect may apply at all, in which case we get pattern ❺.

(1) *Presence and quality of onset interaction in stress*

		Quality of onset	
		YES	NO
Presence of an onset	YES	Pirahã (Arabela) <sup>1</sup> (❶)	Aranda, Banawá, Dutch (❷)
	NO	Karo (Arabela) (❸)	Standard Greek, Russian, etc. (❹)

However, the presence of moraic onsets is not unrestricted. Following Morén (1999/2001), who makes this distinction for moraic codas, I have argued that moraic onsets too are divided into distinctive and coerced ones. The former refer to underlying weight distinctions, while the latter refer to weight acquired in the output as a result of a requirement such as word minimality, weight-by-position etc.

In distinctive weight, the underlying contrast is between /C/ and /C<sup>μ</sup>/, which for the case of onsets, shows up as a contrast between [CV] and [C<sup>μ</sup>V] respectively. Empirically, this translates to a contrast between word-initial singletons and geminates in languages such as Pattani Malay and Trukese (chapter 6). Moreover, I have argued that this contrast can also be found word-medially. To this end, I have provided supporting evidence from Marshallese medial geminates and their relationship to stress and reduplication.

<sup>1</sup> Arabela most likely lacks onsetless syllables, therefore it cannot serve as a testing ground with respect to the presence of the onset issue. This is why I position it in both cells.

In coerced weight, constraints such as WDMIN or MORAIC ONSET<sup>2</sup> render onsets heavy. In this case there is no evidence for an underlying contrast between singletons and geminates, but on the surface some onsets do appear moraic. Unlike distinctive weight which is lexically specified, and consequently potentially arbitrary, in coerced weight, weight is shaped by surface markedness constraints that regulate which types of onsets will end up moraic. In particular, I argue that the markedness consideration relevant to onset weight is voicing and the lack thereof.

Justification of this property as the relevant one for onset weight originates in well-established parallels in tone languages. It is well-known (Kingston and Solnit 1988b, Yip 2002 among others) that voicing becomes relevant in many tone languages, so that voiceless onsets usually raise the pitch of the following vowel, whereas voiced ones lower it. During the course of history - but also synchronically, see Kammu in §1.4.1.1 - some of these languages have lost the voicing contrast, and nowadays uniformly present voiceless onsets on the surface. However, this contrast has not been eliminated entirely; in fact, it has been re-interpreted in terms of tones so that L tones emerge where there used to be a preceding voiced onset and H tones where there was originally a voiceless onset. This pattern then clearly suggests that there exist languages where pitch perturbation due to voicing is interpreted as tone.

What I propose in this work is that in some other languages, namely Pirahã, Arabela and Karo, pitch raising due to the lack of voicing marks stress and is phonologically represented by means of moraic onsets, so that the preferable moraic onsets are the voiceless ones. This is highlighted in the fixed ranking below.

(2) *Onset moraic markedness hierarchy*

$*\mu/\text{ONS}/[\text{voice}] \gg * \mu/\text{ONS}$

In order that any of the onsets actually surface as moraic, a constraint that forces moraicity, i.e. MORAIC, must be interleaved in this hierarchy. This can be placed in any of the following three positions yielding an equal number of different patterns each time.

---

<sup>2</sup> In what follows - in the main text - I abbreviate MORAIC ONSET as MORAIC. Also, recall that MORAIC ONSET, i.e. “onsets are moraic” is the equivalent of WBYP (or MORAIC CODA): “codas are moraic” for onsets. While the introduction of such a constraint may seem stipulatory, I argue that it is no more stipulatory than WBYP whose existence is not justified beyond the fact that it serves well for explaining numerous facts. MORAIC ONSET has the same effect in certain onsets. It would of course be desirable to identify additional supporting evidence for these constraints, but at present, I claim that any objections against MORAIC ONSET on the grounds of being stipulatory are not sufficient, given that this is not a problem inherent to this constraint only.

(3) *Patterns of coerced onset weight*

- a.  $*\mu/\text{ONS}/[\text{voi}] \gg * \mu/\text{ONS} \gg \text{MORAIC}$
- b.  $*\mu/\text{ONS}/[\text{voi}] \gg \text{MORAIC} \gg * \mu/\text{ONS}$
- c.  $\text{MORAIC} \gg * \mu/\text{ONS}/[\text{voi}] \gg * \mu/\text{ONS}$

(3a) is the situation occurring in most languages, namely that no onsets at all are moraic. (3b) expresses what happens in Pirahã, Arabela and Karo, where only voiceless onsets are moraic and voiced onsets are not. The last pattern in (3c) describes the situation where all onsets are moraic. No clear-cut case of this type has been found, although Bella Coola is a possible case. Bella Coola minimal words admit all [CV] words, but ban [V] ones, suggesting that all onsets including the sonorant ones, are moraic. The problem is that in Bella Coola, there are no voiced obstruents<sup>3</sup>, only voiceless ones and sonorants. Under the interpretation that sonorants are specified as [voice], and since sonorants in CV words are moraic too, the ranking in (3c) would be empirically confirmed. However, this is not the sole possibility, because in a different interpretation that I discuss below, Bella Coola would be compatible with (3b).

As we have seen already (cf. §1.4.1.1 for details), tone patterns suggest that in some languages sonorants pattern with the voiced obstruents and in others they do not, suggesting that while they can be specified as [voice], they do not have to be. Consequently, their status with respect to voicing is dual and their specific behaviour is decided on a language-specific basis. The same dual status is evident in sonorants with respect to onset-sensitive stress. In Pirahã and Arabela, sonorants act as [voice] and behave like voiced obstruents, but in Karo, sonorants lack this feature and pattern with the voiceless obstruents. In this view then, it is possible for Bella Coola sonorants to lack a voice feature too and hence offer another instance of (3b). Unfortunately, examination of the current Bella Coola literature has not resolved this issue.

In the next section (§8.2) I will show that the phonological hypothesis that pitch perturbation due to voicing can serve to mark stress is one that can be tested phonetically. After that, I will point out some directions for future investigation focusing on the possibility of a moraic contrast between CV and CCV syllables (§8.3.1), the existence of onset TBUs (§8.3.2) and the need for further language documentation that pays more careful attention to the role of onsets (§8.3.3).

---

<sup>3</sup> To be more accurate, Nater (1979: 171) claims that there are voiced allophones of obstruents, but these only occur intervocalically, thus they could not assist us in identifying the behaviour of sonorants in [CV] words.

## 8.2 Experiment

As we have just seen, one of the main arguments in the thesis is that while pitch perturbation due to voicing is primarily used for tone, it can also be used for stress. During the production of a voiceless consonant, the vocal folds are stiff, which results in pitch raising. While speakers usually utilize this type of  $F_0$  raising for tone, this phonetic cue is present and can be exploited for other purposes too. I suggest that certain languages, e.g. Pirahã, Arabela and Karo, manipulate it for stress. In chapters 1 and 2, I explored how this is done phonologically and formalized it accordingly by means of moraic onsets. However, it is anticipated that this effect could also be tested experimentally. While scrupulous investigation on this matter is required, at present I provide a tentative and exploratory experiment which was conducted for this purpose. It will be shown that it is in accord with the phonological arguments defended in the previous chapters<sup>4</sup>. The aim of the experiment was to test whether listeners would tend to perceive some syllables as more stressed than others depending on the onset's voicing. The hypothesis was that syllables with voiceless onsets would be perceived as more stressed than syllables with voiced onsets. The experiment's structure and results are laid out below.

First, the sequences *pa* and *ba* were elicited from a male British English speaker, who repeated the phrases "Parker, pa, Parker" and "Barker, ba, Barker" a few times each, taking care to leave a space between words. The syllables in question were placed within the carrier phrases to avoid list intonation and to encourage a natural English intonation with the pitch slightly raised for *pa*. The best examples for each of *pa* and *ba* were then selected based on pitch and clarity. The final choice was made with *pa* approximately 7.6 Hz higher than *ba*.

Next, the chosen examples were fed into PRAAT (Boersma and Weenink, version 4.3, 2005). The amplitude, length, and quality of vowel appeared to be already well matched between the two tokens, but differences - that appeared with regard to vowel length only - were eliminated by direct manipulation. These factors were therefore unlikely to affect the perception of stress.

A manipulation of the  $F_0$  of *pa* followed to produce a further token matching the  $F_0$  of *ba*, by lowering the former's contour by 7.6 Hz in its entirety. This gave it a contour which was very similar to the *ba* token. After the manipulation was finished, the

---

<sup>4</sup> Thanks to Donca Steriade and Moira Yip for suggesting this type of experiment. I am grateful to colleagues Eric Carlson, Mary Pearce and especially Stavroula Kousta for valuable help at various stages of the experiment.

sounds were re-synthesized using PSOLA within PRAAT. The resulting sequences were now: *bā*, *pā* (i.e. the same pitch with *bā*) and *pá* (with a pitch 7.6 Hz higher than the other two).

The next step was to combine the tokens to get syllables of alternating voicing. This was done again in PRAAT. The result was 16 tokens in total that ranged between 2-5 syllables, as shown below. I use ‘flat’ to refer to the alternating *pā* and *bā* sequences, i.e. where the pitch is the same, and ‘high’ to refer to alternating sequences of *pá* and *bā*, i.e. where *pá* is of higher pitch.

i) 4 ‘flat pa-initial’ chains of increasing length

- 1) *pābā*
- 2) *pābāpā*
- 3) *pābāpābā*
- 4) *pābāpābāpā*

ii) 4 ‘high pa-initial’ chains of increasing length

- 5) *pábā*
- 6) *pábāpá*
- 7) *pábāpábā*
- 8) *pábāpábāpá*

iii) 4 ‘flat ba-initial’ chains of increasing length

- 9) *bāpā*
- 10) *bāpābā*
- 11) *bāpābāpā*
- 12) *bāpābāpābā*

iv) 4 ‘high ba-initial’ chains of increasing length

- 13) *bāpá*
- 14) *bāpábā*
- 15) *bāpábāpá*
- 16) *bāpábāpábá*

The sequences were then randomized using the “randomized sequences (without duplicates)” tool available at <http://www.random.org/> and were subsequently placed in Powerpoint. This created a single set of 16 tokens. Three more sets of the same tokens were created in the random way described above. All four sets (rounds) were presented

to 6 British English and 2 American English subjects yielding 64 stimuli in total for each subject<sup>5</sup>. The order in which the rounds appeared was also varied. With the exception of two British and one American subject, the rest were at the time graduate linguistics students. All subjects were ignorant of the purpose of the experiment.

The subjects were asked to click on an audio icon to hear the sequence that corresponded to one of the stimuli. Then on a separate sheet of paper they had to circle the syllable(s) - indicated as dots - that sounded more prominent/stressed to them. Thus, on the answer sheet, the sequence ◎●● indicated a three-syllable word, where the first syllable was considered stressed. This procedure was repeated for all stimuli. The subjects were encouraged to note down their first reaction, without thinking about it much. All listeners returned sheets where at least one syllable was circled for each item. A couple of them also seemed to have further intuitions about what sounded as primary and what as secondary stress, but as there were only a handful of such data, this effect was ignored.

Subsequently, for each subject, the percentage of stressed *pa*'s over the total number of stressed syllables in each token was computed. This yields  $16 \times 8 = 128$  averages, which were used in the statistical analysis that follows. To visualize how each of these percentages was produced, consider the following example.

(4) *Average for token [pábā] for subject X*

(N.B: This is a hypothetical example to illustrate the point at hand more clearly. No subject gave this response. Stress is indicated in bold.)

<i>Rounds</i>	<i>Subject's response</i>	<i>Stressed pa's / Stressed əs</i>	<i>Column 3 in %</i>	<i>Average (%) of stressed pá in 4 rounds for the token [pábā]</i>
Round 1	<b>p</b> aba	1/1	100%	62.5 %
Round 2	<b>p</b> aba	0/1	0%	
Round 3	<b>p</b> aba	1/2	50%	
Round 4	<b>p</b> aba	1/1	100%	

Using the averages generated by (4), new averages were produced for each of the 16 stimuli for all 8 subjects as shown below. These, as we will see, form the basis of the graphs presented later on.

<sup>5</sup> For clarity, I use the term 'token' to refer to any of the 16 sequences described, and 'stimuli' to refer to the 64 sequences presented to the subjects.

(5) *Averages of the 16 tokens*

	flat		high	
	pa-initial	ba-initial	pa-initial	ba-initial
<b>2-syllable</b>	89.06	67.19	95.31	79.69
<b>3-syllable</b>	85.42	35.94	98.96	72.92
<b>4-syllable</b>	84.38	60.94	96.88	82.81
<b>5-syllable</b>	79.79	50.63	98.75	68.75

The results were then fed to the statistical package SPSS and a 2x2x4 ANOVA was performed on the data in order to test generality over subjects. The three variables considered were: pitch (2 levels: flat or high), position (2 levels: pa-initial or ba-initial), and number of syllables (4 levels: 2, 3, 4, and 5 syllables). The main effect of pitch was found significant ( $F(1,7)=17.913$ ,  $p=.004$ ), as was the main effect of position ( $F(1,7)=6.853$ ,  $p=.035$ ) and the main effect of syllable count ( $F(3,21)=4.497$ ,  $p=.014$ ). Only one two-way interaction was significant, namely the syllable\*position interaction ( $F(3,21)=3.900$ ,  $p=.023$ ). The remaining two-way interactions, i.e. pitch\*position and pitch\*syllable, as well as the three-way interaction were not significant.

While all these will be discussed next in detail, we can briefly summarize these results as follows. The main effect of pitch confirms that *pa* was perceived as stressed more frequently when the pitch was high rather than flat, thus corroborating the phonological hypothesis pursued in this thesis, namely that the pitch raising offered by the lack of voicing can be construed as stress. The main effect of position also points to the fact that *pa* was perceived as stressed more frequently in pa-initial words than ba-initial words, while the main effect of syllable indicates higher frequency of *pá* in words with an even number of syllables (2 and 4) compared to those with an odd number of syllables (3 and 5). The significant syllable\*position effect, shows that among the ba-initial words, it was in the odd-numbered ones that *pá* was chosen less often.

Furthermore, a one-sample t-test was also conducted to test whether the subjects' selection of stressed *pa* was due to chance or not. For all pa-initial words, regardless of number of syllables and pitch, the t-tests were significant, implying that the selection of *pá* was not a chance effect. On the contrary, for the ba-initial words, with the exception of the high disyllabic and the high quadrisyllabic ones, the test yielded insignificant values, suggesting that *pá* and *bá* could be chosen with the same frequency.

However, since some of the distributions of the data were not normal, the non-parametric sign test was also run<sup>6</sup>. The following results obtained:

(6) *Sign-test results*

Word Type			Significance
<i>Pitch</i>	<i>Position</i>	<i>Number of <math>\sigma</math></i>	
Flat	pa-initial	2	.00781
High	pa-initial	2	.00781
Flat	pa-initial	3	.0156
High	pa-initial	3	.00781
Flat	pa-initial	4	.00781
High	pa-initial	4	.00781
Flat	pa-initial	5	.0156
High	pa-initial	5	.00781
Flat	ba-initial	2	.219
High	ba-initial	2	.0703
Flat	ba-initial	3	.688
High	ba-initial	3	.453
Flat	ba-initial	4	.727
High	ba-initial	4	.125
Flat	ba-initial	5	.727
High	ba-initial	5	.453

These are similar to the results of the t-test, since the selection of stressed *pa* is highly significant in all cases, meaning that its selection was not at all random. The difference with the test above is that the selection of the stressed syllable in *all* ba-initial words has proven insignificant, thus it could be attributed to chance<sup>7</sup>.

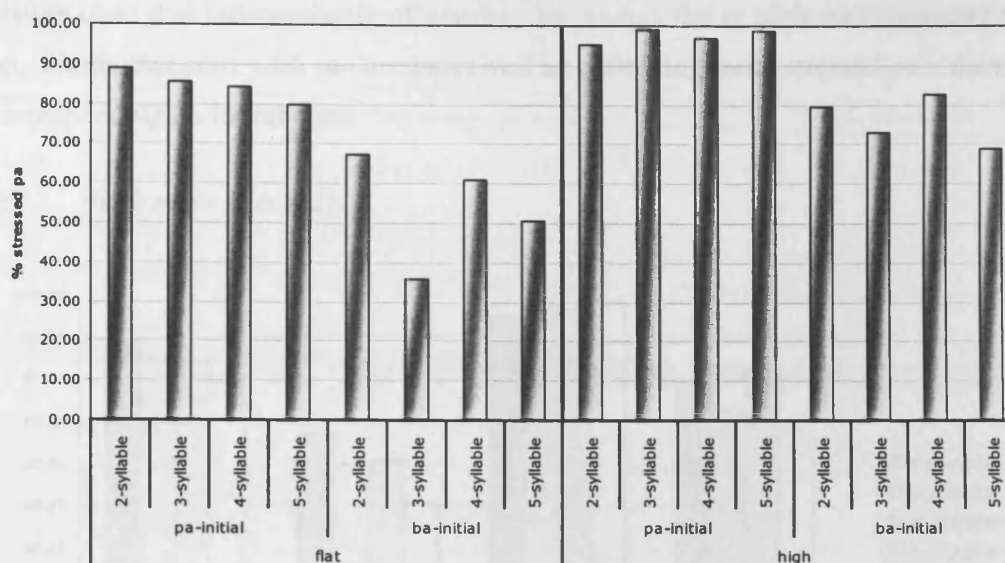
Moving away from statistics, the relevant effects can be visualized more clearly in terms of graphs based on the averages in (5). Since our focus is on pitch, the following graph clearly presents the difference in the behaviour between flat and high pitch.

<sup>6</sup> Available on: [http://fonsg3.let.uva.nl/Service/Statistics/Sign\\_Test.html](http://fonsg3.let.uva.nl/Service/Statistics/Sign_Test.html), retrieved 2 March 2006.

<sup>7</sup> As is discussed below in detail, and highlighted in (5), in all cases, *pa* was stressed more often than *ba*, with the exception of a trisyllabic flat ba-initial word where the reverse occurs, and in a pentasyllabic flat ba-initial one where *pá* and *bá* are roughly equally chosen. This is why in this paragraph in the text, for pa-initial words, I refer to stressed *pa* and for ba-initial ones, I refer to the stressed syllable, i.e. the more general term, given that there both *bá* and *pá* occur.



(7) *Pitch main effect (1)*

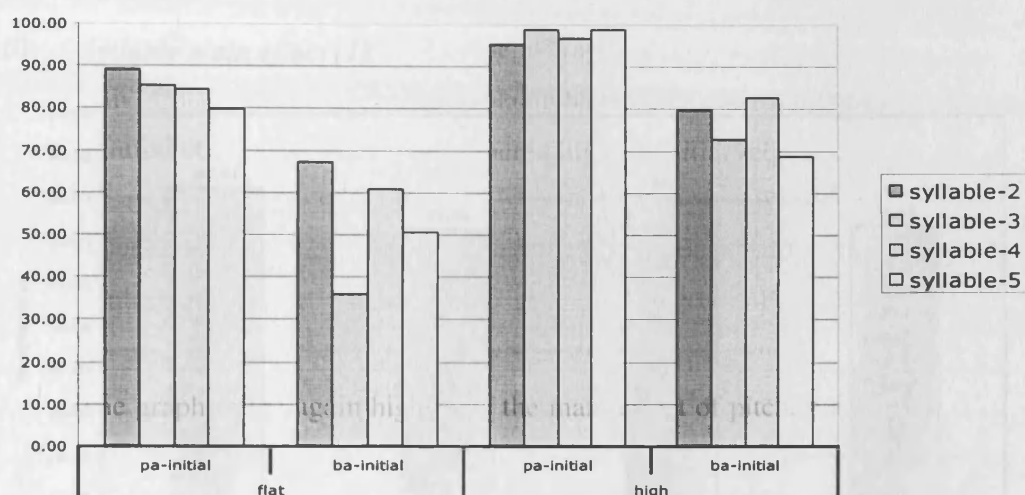


This graph can be split into two large areas (indicated by the solid line in the middle). The one on the left presents the results of stressed *pa*'s in flat chains, while the one on the right does the same for high chains. It is clear that while listeners tend to stress *pa* much more often than *ba* overall, this effect is greatly exaggerated when the pitch is also higher. This conforms with the phonological argument made before, namely that the pitch raising caused to the vowel after a voiceless onset can be interpreted as stress. Of course, one may note that in the flat chain too - where pitch is the same between *pa* and *ba* - listeners still perceive syllables with the voiceless onset as stressed, although it could be conceivable that stressed *pa* would be chosen only half of the time, so that the average values in the flat chains should be in the area of 50%. While this is a valid point, it is also the case that the pitch perturbation due to a voiceless consonant would be hard to suppress entirely (Yi Xu, p.c.), thus it is possible that subjects carry over this effect even in the cases where the pitch of *pa* is equal to the one of *ba*.

However, the crucial point seems to be that increased pitch in the high chain acts as a major disambiguation factor for stress. What is meant by that is that among the instances of flat chains, there are a few where either *ba* is perceived as stressed more often than *pa* (i.e. in flat trisyllabic *ba*-initial words) or where the selection of the stressed syllable is close to being random (i.e. in flat pentasyllabic *ba*-initial words). In the high chains, no such effect arises. The averages for stressed *pa* selection range from a minimum of 68.75% in a high pentasyllabic *ba*-initial word to values that nearly reach the maximum 100%, as in e.g. odd-numbered high *pa*-initial sequences.

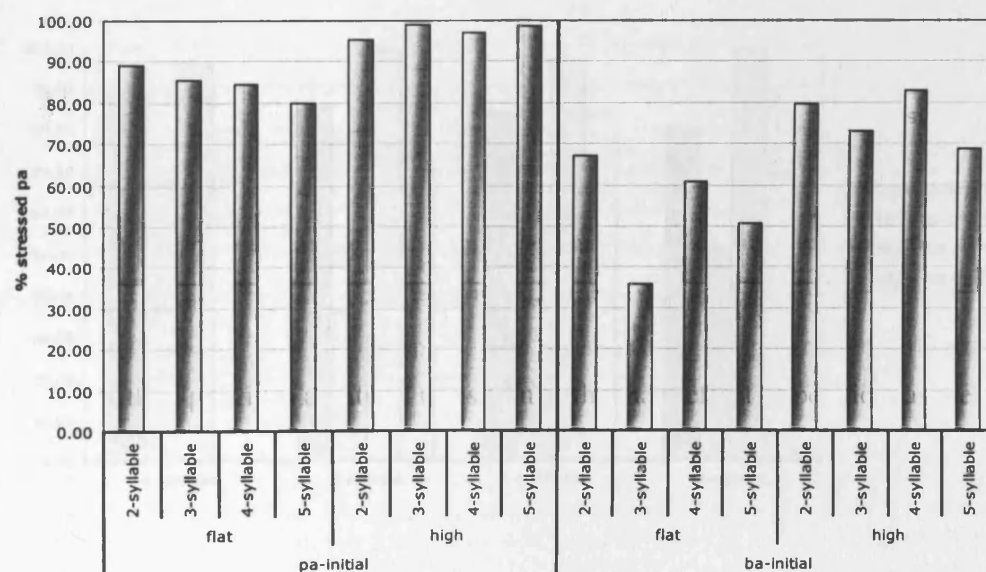
The graph in (8) again highlights the main effect of pitch, but this time it is also visibly clear that independently of whether the pitch is flat or high, an asymmetry crops up. Words that start with *pa-* are perceived as containing more stressed *pa*'s than their corresponding *ba*-initial ones.

(8) *Pitch main effect (2)*



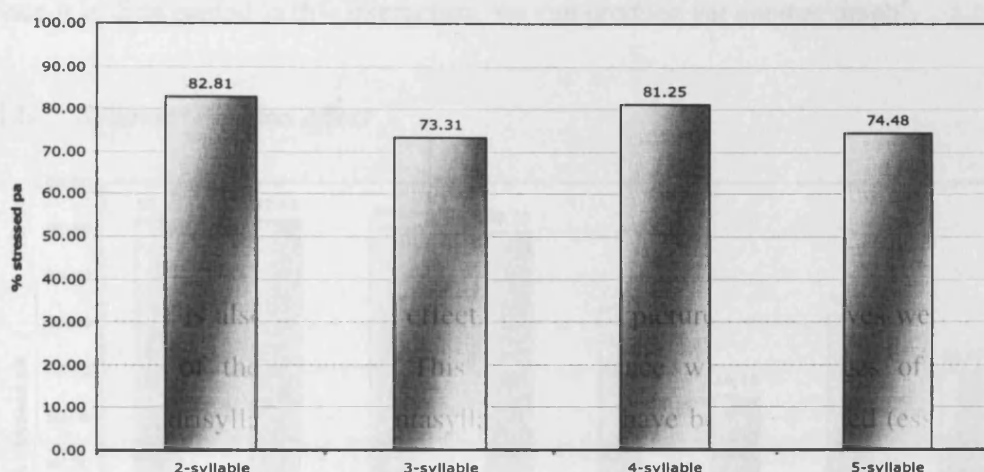
This graph then provides a natural transition to the main effect of position, as mentioned above. Again if we divide (9) in two large chunks, the leftmost represents the percentages of stressed *pa*'s in *pa*-initial words, and the rightmost shows the same effect in *ba*-initial words. Evidently, the former systematically show more stressed *pa*'s than the latter.

(9) *Position main effect*



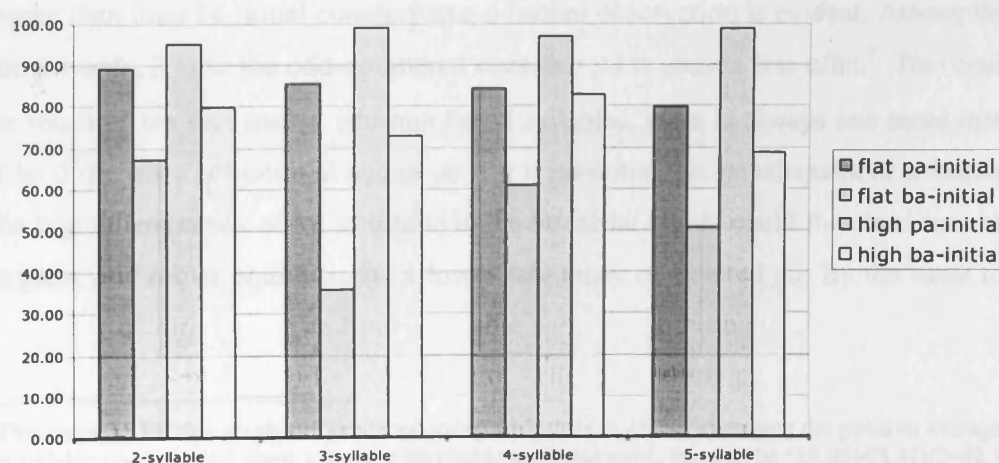
However, there is also a syllable effect. The rough picture below serves well as a first approximation of the situation. This is an instance where averages of disyllabic, trisyllabic, quadrisyllabic and pentasyllabic words have been computed (essentially by averaging each row in (5) for each type of syllable ignoring pitch and position). Even-numbered syllables present a higher percentage of stressed *pa*'s compared to the odd-numbered ones.

(10) *Syllable main effect (1)*



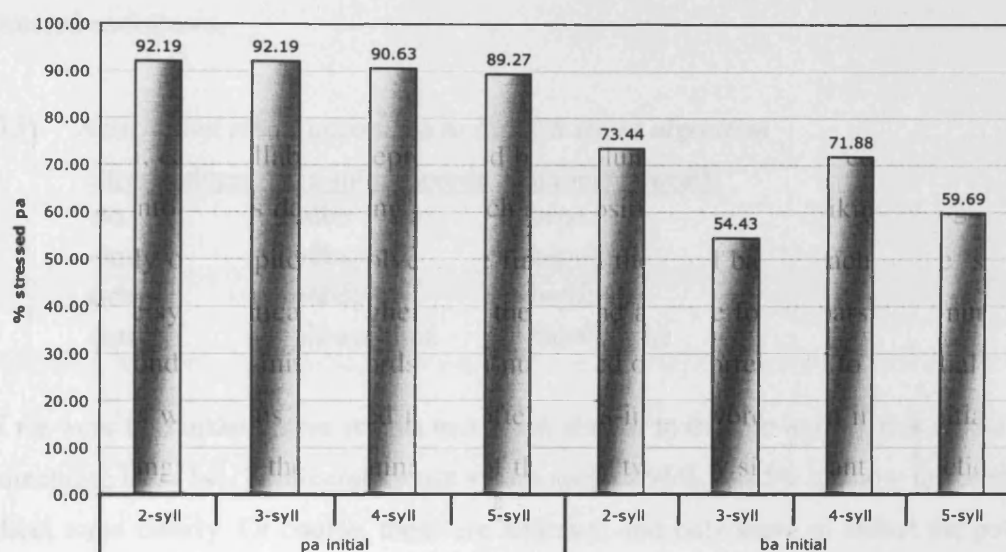
The graph below re-iterates this result in a less obvious way, but by breaking down the syllable effect according to the pitch and position involved, it provides information about the correlation between syllable and position.

(11) *Syllable main effect (2)*



In this graph, each syllable is represented by a column (the 'syllable column') which is subdivided into 4 bars depending on pitch and position. What is striking though is that independently of the pitch involved, the first and the third bar of each of the 'syllable columns' are systematically higher than the second and the fourth bars. Odd-numbered bars correspond to *pa*-initial words, even-numbered ones correspond to *ba*-initial words. Stressed *pa*'s were thus selected more often in *pa*-initial words than in *ba*-initial ones. This then brings us to the examination of the only two-way significant interaction, that of syllable\*position mentioned earlier on ( $F(3,21)=3.900$ ,  $p=.023$ ). Setting aside pitch, since it is disregarded in this interaction, we can produce yet another graph<sup>8</sup>.

(12) *Syllable\*Position effect*



While we can clearly see that stressed *pa*'s are chosen in *pa*-initial words to a higher degree than their *ba*-initial counterparts, a further observation is evident. Among the *ba*-initial words, it is in the odd-numbered ones that *pá* is chosen less often<sup>9</sup>. This could be the result of the fact that in odd-numbered syllables, there is always one more instance of *ba* if the word is *ba*-initial and of *pa* if it is *pa*-initial, i.e. *bapabapaba* or *pabapabapa*. The higher frequency of *ba* in odd-syllable *ba*-initial words could thus lead to a higher frequency of *bá*, or equivalently, a lower frequency of stressed *pá*. By the same token,

<sup>8</sup> The averages for this graph are produced using the values in (5) by averaging the position average with the syllable one ignoring pitch, e.g. for a disyllabic *pa*-initial word, this will be  $(89.06+95.31)/2=92.19$ .

<sup>9</sup> In a later examination of the tokens, it was found that the first *ba* in the flat *ba*-initial odd-numbered words, i.e. in flat *bapaba* and *bapabapaba*, was actually slightly higher in pitch, than the normal flat *ba*. It is possible that this has affected the results to a certain extent. Unfortunately, due to time pressures, it was impossible to re-run (this part of) the experiment with the problem corrected.

we would expect that more stressed *pa*'s should appear in odd-syllable *pa*-initial words, and yet this does not occur.

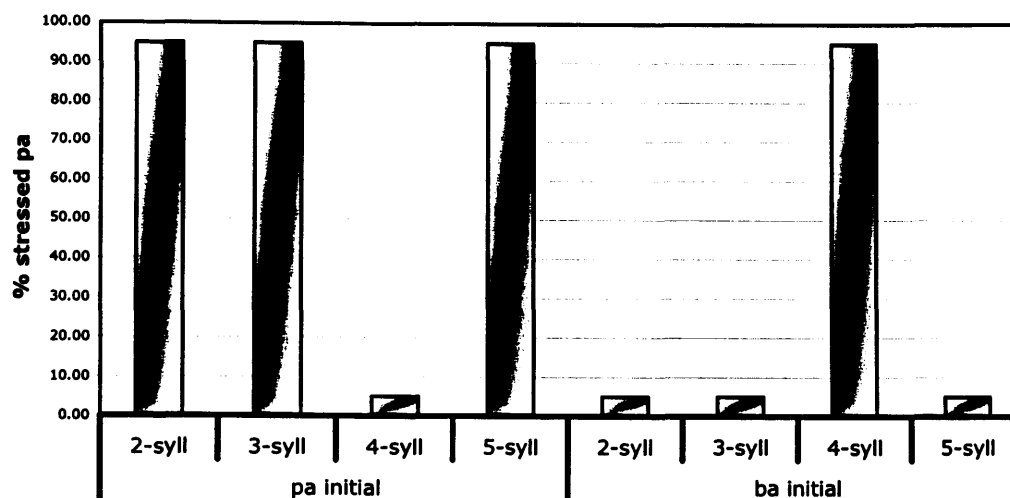
These effects are particularly interesting if we consider English stress in general. While English stress is notoriously complex (Halle and Keyser 1971, Hammond 1999), some patterns are rather systematic. Since presumably the sequences considered in this experiment would be treated as nouns, then the general algorithm of the language would dictate that disyllabic words receive initial stress, e.g. [hæpi] 'happy', [óprə] 'opera', [máni] 'money', and longer words would get antepenultimate primary stress, by virtue of the fact that all syllables are light and open, e.g. [pærədi] 'parody', [əmérikə] 'America' and perhaps secondary stress on the initial syllable if it is long enough, e.g. [prəʊlɪgómɪnən] 'prolegomenon'. Schematically then, the examples used here would be stressed as follows.

(13) *Anticipated stress according to English stress algorithm*

<u>Stress pattern</u>	<u>pa-initial words</u>	<u>ba-initial words</u>
σσ	pába	bápa
σσσ	pábapa	bápaba
σσσσ	pabápaba	bapábapa
σσσσσ	pàbapábapa	bàpabápaba

If we were to translate these results to a graph similar to the one in (12), this would look something like (14). I deliberately use values such as 95% and 5% to show the predicted effect more clearly. Of course, these are arbitrary, and only serve to reflect the point in question. In all *pa*-initial words, with the exception of a quadrisyllable, all stressed syllables should be *pá*'s (high % of *pá*). In a quadrisyllable, they should be *bá*'s, hence an extremely low percentage of *pá*. The mirror image obtains for *ba*-initial words.

(14) *Anticipated appearance of (12) under normal English stress*



If we compare (12) with (14), we can observe that the frequency of stressed *pa*'s in the *pa*-initial words of (12) matches the one in (14) pretty accurately, with the exception of quadrisyllabic words where it is markedly different. There (14) predicts very low frequency of stressed *pá*, while the facts in (12) show very high frequency.

For *ba*-initial words, the results are much more perplexing. (14) predicts very low frequency of stressed *pa*'s for all cases with the exception of quadrisyllabic words, where *pá*'s should be highly frequent. In reality, (12) shows that overall, the frequency of stressed *pa*'s is lower than the one compared to the *pa*-initial words, but still quite high. As mentioned, in fact there is a syllable\*position effect with odd-syllable *ba*-initial words presenting lower frequency of *pá*'s compared to even-syllable ones. In (14), no such effect emerges; more astonishingly (14) actually predicts that even-syllable *ba*-initial words not only are not grouped together, but that they have a dramatically different behaviour: stressed *pa*'s in disyllables should be scarce, but in quadrisyllables, they should be abundant.

Although as we have seen there is a rather good match between (12) and (14) in the behaviour of *pa*-initial words, this is by no means the case in *ba*-initial words. While this discrepancy is puzzling, it is important to note that the results of the experiment are really interesting, as they highlight that the normal English stress algorithm is insufficient to account for the attained results. Obviously, the difference in voicing between *pa*'s and *ba*'s must be one of the major factors, especially with respect to *ba*-initial words, which despite the expectations in (14) present a rather high frequency of stressed *pa*'s in words of all sizes. The remaining effects require further investigation.

To sum up this section, I have shown that the experimental results confirm that *pa* was perceived as stressed more frequently in the high rather than the flat conditions, a fact that verifies the phonological hypothesis pursued throughout; *pa* was also perceived as stressed more frequently in *pa*-initial words and in words with an even number of syllables (2 and 4) rather than words with an odd number of syllables (3 and 5). There is also a syllable\*position effect, where among the *ba*-initial words, it is in the odd-numbered ones that *pá* is chosen less often.

Of course, this experiment generates new research questions. Obviously, the next thing to do would be to conduct an experiment along these lines in Pirahã, Karo and Arabela and see what the effect of pitch due to voicing is on stress. Although these languages present other complications too, such as the independent existence of tone in Pirahã and Karo and strong right edge effects, e.g. trisyllabic window (Pirahã) or preferences for rightmost stress (Karo and Arabela), we would foresee that the results should be similar, and in fact maybe even further exaggerated given that these languages base (part of) their stress algorithm on voice-conditioned pitch.

Future experiments would also need to add sonorants into the equation and see whether the hypothesis about their ambiguous nature is verified. In languages like Karo, they should induce similar results with voiceless obstruents, but in others, such as Pirahã and Arabela, they should pattern with the voiced obstruents.

Moreover, it would be interesting to see at what point the effect of voice-conditioned pitch is actually integrated into the stress algorithm of the language. In other words, if it is cross-linguistically true that pitch raising due to the lack of voicing can be interpreted as stress, then when does this become a key element of stress assignment, so that a language like English - which only shows these effects under controlled experimental conditions - also presents them in natural speech as in Pirahã? Further research is thus definitely required to shed light on these and other questions that spring from the interesting interaction between stress and onsets, hopefully offering a better understanding of the phenomenon of stress as a whole.

### **8.3 Directions for future research**

As we have seen, the introduction of moraic onsets and their particular distribution conditioned by voicing, accounts for a number of languages which had been explored in the preceding chapters, but also proposes a more general representation for some phenomena, e.g. initial geminates as moraic onsets (particularly useful for Austronesian

languages which have them in abundance). Additionally, it makes predictions about the patterns we should anticipate with respect to onset weight and voicing, i.e. if an onset is moraic, then it should always be the case that the voiceless consonants are weightful.

This system also brings other considerations to the surface, all of which merit further research. Here I will address three: a) the possible moraic contrast between CCV and CV (§8.3.1), b) onsets as tone-bearing units (§8.3.2) and c) (additional) documentation of existing and new data (§8.3.3).

### **8.3.1 Moraic contrast between CV and CCV?**

The first issue posits the question of whether it is possible to have a language where complex onsets are weightful, but singletons are not, i.e.  $CCV > (C)V$ . At present, I can only offer some speculations on this matter. First, recall from Chapter 7 that two languages which were thought to have exactly this system according to Gordon (2005), namely Nankina and Bislama, were re-considered in the light of fuller examination of the languages' phonology. It was concluded that this assertion was only superficial and no relevant convincing argument could be constructed.

Furthermore, in Chapter 5, I proposed an account of Samothraki Greek compensatory lengthening that made no use of onset weight. As I had argued then, such an analysis was enforced in order that basic tenets of OT, such as the Richness of the Base and its fully parallelistic nature, were observed. However, an analysis which is not bound by such considerations, e.g. a standard serialist CL analysis along the lines of Hayes (1989) with the addition of moraic onsets, would also be possible. The fact that SamG /r/ causes CL of the following vowel in onset consonant clusters, could suggest the existence of  $CrV > CV$  that is comparable to the sought-after  $CCV > (C)V$ . Nonetheless, /r/ also appears moraic as a singleton (word-initially), thus one would probably need to say that /r/ is underlyingly moraic regardless of whether it appears in a complex or singleton onset. In this view, the  $CCV > (C)V$  idea no longer could be maintained, since both  $CrV$  and  $rV$  would behave in the same way.

Data from another language that could support the  $CCV > (C)V$  contrast are provided by Luganda metrics (§7.2.4.1), where syllables with a complex onset whose second consonant is a glide and - potentially - syllables with a full or partial geminate count as heavy in poetic meter. As I had highlighted in that section, this case is not unambiguous either, since adopting suggestions like Smith's (2003) 'nuclear onglides'



and the possibility of resyllabification sidestep the need to consider CCV syllables as heavy. It is thus an open issue whether the contrast  $CCV > (C)V$  can ever arise.

Note that a similar contrast does not arise in codas either, that is, I am not aware of any language where:  $(C)VCC > (C)VC, (C)V$ . This is a language that would treat open syllables with a short vowel and those with a singleton coda as light, but once the coda would acquire an extra consonant, then this would become heavy. Although  $(C)VCC > (C)VC > (C)V$  is attested, as in Hindi (Hayes 1995), where  $(C)VCC$  syllables are superheavy,  $(C)VC$  ones are heavy and  $(C)V$  ones light, the previous weight scale is absent. Presumably, the absence of such patterns can be attributed to the fact that MORAIC ONSET / CODA is really the constraint responsible for moraicity on onsets and codas respectively. Since this merely looks at these syllable positions without distinguishing between singleton and complex margins, then it would be impossible that a complex onset / coda becomes moraic, without the singleton becoming moraic too. Thus, the implication arising is that if a language has weightful complex syllable margins, then it should also have weightful singleton margins too. In principle then, a language that possesses the following system with respect to onset weight:  $CCV > CV > V$  is possible. I have not been able to find such a language, but as it is a predicted one, future investigation will perhaps be able to verify it empirically.

### **8.3.2 Onsets as TBUs?**

The second issue I address here is concerned with tone. Since the topic of tone and tone-consonant interactions is complex and far from straightforward, as the vast literature on it (e.g. Halle and Stevens 1971, Yip 1980, 1995, Bao 1990, Duanmu 1990, Peng 1992, Bradshaw 1999 and many others) suggests, at present I can only touch upon it and make some general statements, which should merely serve as a point of departure for further research.

With this disclaimer in mind, the point of interest here is the following; since I have argued that moraic onsets exist, and since moras serve as tone bearing units (TBUs) in several languages, a natural expectation would be to find moraic onsets that carry tone. I will argue that while this is possible, there are good reasons why it is vanishingly rare. Despite that, still it is attested, as data from at least one language, namely Kpelle, a Mande language spoken in Liberia, suggest.

(15) *Kpelle onsets as TBUs*

Hyman (1985: 44): initial obstruents and sonorants

a.	<i>Stem</i>	<i>'my'</i>	<i>'his/her'</i>	
i.	pólù	míbólù	b́ólù	'back'
	túé	ndúé	dúé	'front'
	kśś	ńgśś	ğśś	'foot, leg'
	fíí	ḿv́íí	v́íí	'hard breathing'
ii.	lēē	ñēē	hēē	'mother'
	yéé	j́éé	h́éé	'hand, arm'
	mā́lón	ḿā́lón	m̃ā́lón	'misery'
	ńń	j́ńń	h́ńń	'tooth'
	Welmers (1962: 72): intervocalic obstruents			
b.	kāpa	'penny'	bêbe	'raffia purse'
	ğbete	'fix it'	ğbôdo	'leprosy'

In the first set of data (a), we can see that for the production of the 'my' form, a high-toned nasal is prefixed to the stem. If the stem happens to start with an obstruent, then voicing assimilation also occurs (a.i), but if it begins with a sonorant, then total assimilation and nasal simplification occurs (a.ii) [N.B.: I am glossing over the specifics of these processes, as my focus here is on tone and onsets]. The 'his/her' forms appear with either an initial low-toned voiced obstruent or with a low-toned nasal. In the light of the above, and due to the existence of minimal pairs such as (16), where a sonorant onset can appear toneless, L-toned or H-toned, Welmers proposes that in the *stem* - 'his/her' forms of (15a), the underlying distinction for [n]~[h̃] is one between /n/~/̃ n/, while for [p]~[b] word-initially, it is one between /p/~/̃ p/.

(16)	mare-kêi 'a question'	m̃are ké 'ask him'	ḿare ké 'ask me'
------	-----------------------	--------------------	-------------------

Moreover, Welmers proposes that the word-medial voicing contrast for obstruents shown in (15b), provides evidence that there is also an underlying /b/ that occurs word-medially. His proposal can thus be summarized as:

(17) *Phonemes and allophones according to Welmers*

(a) /p/			
	/p/	→	[p] word-initially, e.g. pólù
	/̃ p/	→	[b̃] word-initially, e.g. b́ólù
(b) /b/			
	/b/	→	[b] intervocalically, e.g. bêbe

However, I would like to re-interpret these contrasts in a slightly different way so that actually there is only one phonemic obstruent, and all other ones are derived allophonically. This can be supported by at least two pieces of evidence. First, Welmers himself notes that the initial [b] is heavily voiced and begins with low pitch, which is consistent with the fact that it derives from underlying [<sup>ˀ</sup>p] with a floating low tone. On the other hand, the intervocalic [b] is not heavily voiced and does not begin with a low pitch. This could suggest that in fact it is not underlyingly voiced, but it could simply derive from plain voiceless and toneless /p/, which shows up as slightly voiced due to ambient voicing. Second, Welmers does not fully discuss what happens with the underlying representations of the sonorants presented in (16). Following his reasoning in (17a), the obvious thing to say would be that [m] comes from /m/, [m̃] from /<sup>ˀ</sup>m/ and [m̂] from /<sup>ˈ</sup>m/. Adding all these together, we get:

(18) *Phonemes in Kpelle and their allophones --- present proposal*

- |     |                      |   |   |
|-----|----------------------|---|---|
| (a) | Obstruents, e.g. /p/ |   |   |
|     | / <sup>ˀ</sup> p/    | → | [p] word-initially, e.g. pólù                   |
|     | / <sup>ˈ</sup> p/    | → | [b̃] word-initially, e.g. b̃ólù                 |
|     | /p/                  | → | [p] intervocalically, e.g. b̃êb̃e               |
| (b) | Sonorants, e.g. /m/  |   |   |
|     | / <sup>ˀ</sup> m/    | → | [m̂] word-initially, e.g. m̂are                 |
|     | / <sup>ˈ</sup> m/    | → | [m̃] word-initially, e.g. m̃àr̃e                |
|     | /m/                  | → | [m] word-initially/medially, e.g. mar̃e / dàmaa |

The difference with Welmers is that while he imposes two phonemes for surface [b], I only propose one, namely /p/, which depending on the presence of floating tone can receive various manifestations. The second point is that I also allow for an underlyingly H-toned obstruent - in analogy to sonorants - whose tone however fails to surface, because, as I will argue next, voiceless obstruents cannot be surface TBUs.

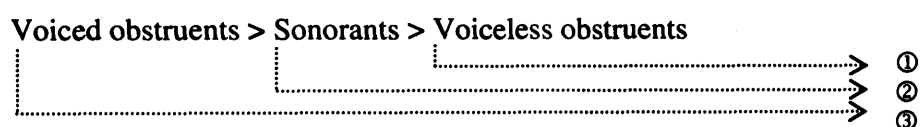
Assuming this is right, then we would need to say that this high tone either deletes, remains afloat or shifts to a neighbouring segment. The final possibility, although appealing, is very difficult to test, since there is no agreement among researchers on some of the facts or the exact tonal association. For instance, while Welmers (1962: 75) transcribes the stem of the word meaning ‘back’ as [pôlu] with a contour tone, Hyman (1985: 44) gives it as [pólù]. As Hyman also acknowledges, the forms in (15) have caused a “heated debate” over the correct transcription of the facts and their analysis, thus it would take us too far afield to provide a fuller analysis at this stage. Notably though, the fact that certain onsets in Kpelle appear as tone-bearing is

undisputable. This can naturally be accommodated in a framework that admits moraic onsets, which consequently can act as TBUs.

Two theoretical issues now need to be addressed. First, why is the phenomenon of onsets as TBUs so remarkably rare? I will argue that this is because TBUs and moraic onsets pose conflicting demands. And, second, why have I claimed that voiceless obstruents cannot be TBUs? This is because voiceless obstruents do not satisfy the phonetic conditions required for TBUs. The remainder of this section elaborates on these ideas. I will start the discussion by considering moraic onsets - the presence of whom is quite rare already - and what constitutes the best moraic onset.

As I have argued, the preferred moraic onsets are the voiceless obstruents due to the fixed ranking:  $*\mu/\text{ONS}/[\text{voi}] \gg * \mu/\text{ONS}$ . Thus, if a language only possesses a single type of moraic onsets, then these will be the voiceless obstruents (19.①), as in Pirahã and Arabela due to  $*\mu/\text{ONS}/[\text{voi}] \gg \text{MORAIC} \gg * \mu/\text{ONS}$ . The same ranking, but with the added proviso that sonorants lack the feature [voice], generates a language where both voiceless obstruents and sonorants are moraic onsets (19.②), like in Karo. Finally, the ranking  $\text{MORAIC} \gg * \mu/\text{ONS}/[\text{voi}] \gg * \mu/\text{ONS}$  entails that all onsets will be moraic, i.e. voiceless obstruents, sonorants and voiced obstruents (19.③). Bella Coola is perhaps an example of this sort with respect to Word Minimality as shown in Chapter 4. There is thus an implicational relationship, depicted below.

(19) *Implicational relationship for moraic onsets*



Consider what the consequences of these patterns are for tone. Since the physical correlate of tone is fundamental frequency, only voiced segments should be able to carry it, i.e. vowels, sonorants and voiced obstruents<sup>10</sup>. While the harmonics of voiced obstruents are low in energy compared to those of the sonorants, they could still be expected to participate in tonal assignment, albeit to a very limited extent (Gordon 1999). It thus follows that while a language may be diagnosed through weight-based

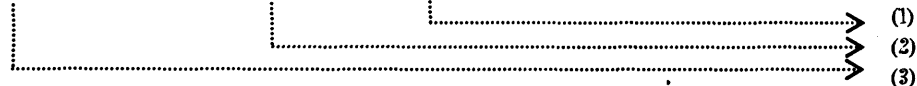
<sup>10</sup> Note that the use of 'voiced' here does not necessarily correspond to phonological [voice], since sonorants, which can carry tone, may lack this feature. 'Voiced' in this part of the discussion, merely refers to the presence of fundamental frequency and harmonics. The fact that a voiceless segment itself lacks  $F_0$  does not of course imply that it cannot affect the  $F_0$  of neighbouring segments, as we have seen in the interaction between pitch and tone (and stress) in §1.4.1.1.

phenomena to treat all segments - including obstruents - as moraic, only a subset of these may actually be TBUs, an idea originating in Steriade (1990).

To see this more clearly, let us consider languages where  $CVV=CVC=2\mu$ , i.e. where the coda is moraic, and focus on what the TBU can be. In many languages (20.1), only vowels in nuclei receive tones as in Ancient Greek (Steriade 1990). In others, such as Danish (Zec 1988, Steriade 1990), Early Greek, (roughly equivalent to Homeric Greek; Steriade 1990), Lithuanian (Zec 1988, Steriade 1990, Gordon 1999), Kiowa (Gordon 1999) and Kunama (Connell, Hayward and Ashkaba 2000), the coda in CVC syllables is moraic no matter its quality, but only sonorant codas can bear tone alongside the vocalic TBUs of nuclei (20.2). Finally, in Hausa, Musey and Luganda, Gordon (1999) finds that not only sonorants and vowels are TBUs, but also the much rarer (voiced) obstruents (20.3)<sup>11</sup>. The picture thus shaped is presented below.

(20) *Implicational relationship for TBUs*

- a. Voiced obstruents > Sonorants > Vowels



- b. Voiceless obstruents are not (surface??) TBUs universally<sup>12</sup>

Now, the idea is that, since voiceless obstruents cannot be TBUs, it should be impossible to find tone on moraic onsets in a language with pattern (19.①) where only voiceless moraic onsets are admitted. Recall that since (19.①) is the default instance of moraic onsets, by eliminating this pattern, the likelihood of moraic onsets as TBUs is automatically *hugely* reduced. We are thus left with patterns (19.②) and (19.③). In (19.②), both sonorants and voiceless obstruents are moraic, but due to the inability of the latter to be TBUs, only the former can assume this role. Thus, we can predict that it should be possible for a language to have the moraic onset pattern of (19.②) and the TBU pattern of (20.2), where only nuclei and moraic sonorant onsets receive tone. In (19.③), where all segments can be moraic onsets, the prediction is that TBUs can either be of the type in (20.2), i.e. nuclei and sonorant onsets or of the type in (20.3), where

<sup>11</sup> I discuss Musey in more detail below. As for the other two, Hausa admits contour tones on CVC syllables with obstruents, e.g. *rās:á:* 'branches' (Gordon 1999), which, like Musey, could suggest that voiceless obstruents *phonologically* act as TBUs. In Luganda (Ashton et al. 1954), voiced and voiceless geminates seem able to carry tone, but data are too sparse to reach any conclusion. Notably these are characterized as syllabic, so perhaps an analysis along the lines of sesqui-syllabic languages is in order.

<sup>12</sup> I address this issue next.

apart from vowels in nuclei, also sonorant and voiced obstruent onsets can be TBUs. Kpelle above seems an instance of the latter pattern (Hyman 1985, Welmers 1962).

However, recall that in Kpelle I had argued that voiceless obstruents underlyingly may carry tone (18a), although they might surface as toneless. This relates to (20b), where I question the universality of the ban of voiceless obstruents as TBUs.

To this end, consider Musey (Shryock 1995), cited in Gordon (1999), where consonants are divided into Type A (or High consonants) and Type B (or Low consonants). In the absence of lexical tone, word-initial Type A consonants induce mid tone on the first vowel, whereas Type B trigger low tone. Roughly, Type A consonants include the sonorants and the obstruents that historically used to be voiceless. Type B correspond to the historically voiced consonants. Processes such as rightward displacement of lexical L tone (21), suggest that the contrast between Type A and B consonants is genuine and causes tonal differences. If the mid tone after Type A consonants is treated as being introduced by them, then under the assumption that tones attach to moras, we would need to say that voiceless obstruents were at least at some point phonologically TBUs, although phonetically they are not (since the tone surfaces on the neighbouring vowel). Alternatively, as Shryock suggests, the mid tone can be assigned by default. But in this view too, something special has to be said about why only Type A consonants cause tonal displacement, i.e. why we do not get \*fūnà too.

(21) *Rightward displacement of lexical L tone in Musey*

a.	<i>Cliticisation of /-na/</i>		
Type A:	sà → sanà → sǎnà	'person'	
Type B:	fù → fùnà	'goat'	
b.	<i>Subjunctive</i>	<i>Subjunctive with affixation</i>	
Type A:	tò 'sweep'	tôm	'sweep it'
Type B:	dò 'pick'	dòm	'pick it'

This then raises the issue about the possibility of distinguishing phonological from phonetic TBUs. If this proves necessary, then perhaps voiceless obstruents can be phonological TBUs, albeit not phonetic ones. This latter property explains why Kpelle /p/ is underlyingly admitted, although it fails to surface as such.

Things though are further complicated in the so-called sesqui-syllabic languages (Sloan 1988, Gafos 1998, Lin 1998, Hendricks 2001) like Kammu (Svantesson 1983). In those languages, many words consist of a full syllable and a 'minor syllable' that usually precedes it. A minor syllable typically consists of an onset and a coda. The latter

serves as the syllabic element which, in some sesqui-syllabic languages, can also carry tone. Commonly the coda includes a sonorant, e.g. *tm.ràʔ* ‘stove’, *kr.làn* ‘bulge’, but it can also contain a voiceless obstruent, which despite expectations, can nonetheless carry tone, e.g. *pk.ték* ‘to tell a riddle’, *ks.rìs* ‘to shake something out’. In the light of these facts, it can be argued that the universal ban on surface voiceless obstruent TBUs does not hold either.

However, the status of minor syllables remains elusive (see Lin 1998 for a brief overview of proposed accounts), and their special behaviour is unquestionable, given that they are prosodically defective, they occur at word edges, they cannot bear main stress, they can be insensitive to stress assignment, etc. In fact under one interpretation of minor syllables, the second consonant can be considered the coda of a syllable with an empty moraic nucleus, i.e. [*ks.rìs*] from above is actually [*kʰs.rìs*] (Gafos 1998 on Temiar) or the coda of a reduced vowel, i.e. [*kás.rìs*] (Coleman 1996 on Tashlhiyt Berber). If any of these is on the right track, then the (empty) nucleus, instead of the - sometimes voiceless obstruent - coda, would carry the tone, thus eliminating the problem above. I will thus adopt a position along these lines, and maintain that voiceless obstruents cannot be surface TBUs universally.

In sum, it is anticipated that moraic onsets as TBUs should be an extremely rare phenomenon, given that onset moraicity and tone essentially impose contradictory requirements. The preferable moraic onsets are the voiceless obstruents which lack  $F_0$ , whereas tone needs to dock on segments with fundamental frequency. Thus, moraic onsets as TBUs can only arise in highly-restricted environments, where several conditions are satisfied at once<sup>13</sup>. In this view, their extreme rarity falls out naturally.

---

<sup>13</sup> We can explore one more theoretical possibility, which empirically would be extremely difficult to verify given that languages tend to avoid the simultaneous presence of moraic onsets and codas. However, such cases exceptionally occur as in Karo (§2.4.3) which treats sonorant codas as moraic, and sonorant and voiceless obstruent onsets as moraic too. In the light of such data, we could imagine a language with moraic onsets and codas, and where tones could be admitted on margins. What would this look like? Essentially, despite possible differences in what can constitute a moraic onset and a moraic coda, the crucial point is that TBUs in both positions should be the same, i.e. voiced obstruents and sonorants, thus also frequently yielding symmetric patterns too, e.g. if sonorants are allowed to be TBUs in codas, they could also be TBUs in onsets. Assuming furthermore that a language is found where all types of onsets and all types of codas are moraic, then also voiced obstruents could be TBUs. What is impossible however, is to find a language where for codas both sonorants and voiced obstruents are TBUs, but in onsets, only voiced obstruents are TBUs. This cannot happen due to the hierarchy of (20). If voiced obstruent onsets are TBUs, so must sonorants be.

### 8.3.3 Language documentation and attention to the role of onsets

A final issue springs from the languages discussed in Chapter 7, for which some indication of onset moraicity was available, but in many cases insufficiently supported by the available data. What this suggests is that there exists a pressing need to document and re-examine more languages in the light of the proposal that moraic onsets are present. I contend that a fair number of cases which have been poorly understood as well as other data yet to be studied, may receive satisfactory accounts once the theoretical bias against weightful onsets is removed.

In Chapter 7, I discussed some potential instances of onset sensitivity for which no adequate evidence yet exists. I am however aware of other languages with similar patterns, whose data are even murkier or sparse, and thus disallow us to draw any conclusions. While this is not meant as an exhaustive list, I include these for completeness: Kaxuyâna, Eastern Popoloca, Tümpisa Shoshone and Gadsup stress, as well as various cases of compensatory lengthening mentioned in Rialland (1993).

In Kaxuyâna (Paula 1980), Gordon (1999: 257) claims that the minimal word is CCV or CVCV. However this seems wrong, as it is explicitly stated numerous times that CCV is the product of the first V deletion in CVCV as a result of rapid speech (cf. Paula 1980: 23-24, 55). However, Kaxuyâna could have some different type of onset sensitivity. Spike Gildea (p.c.) mentions that Paula's description of the stress system is largely misguided, and instead in his own field work, Kaxuyâna can clearly be shown to construct moraic iambic feet and count CVC as heavy. The interesting thing is that some of the words that began with CVCV underwent reduction, so that a word-initial CVrV... > CrV... (stress in bold). Now if feet are moraic iambs, stress on CrV should not be expected unless this syllable is considered heavy. Had it been light, then it should form the foot tail of an iamb whose head is the syllable after CrV. Unfortunately, Gildea does not have any data in his corpus that verify this, so not much more can be said about this pattern at present. Note however that if this is indeed attested, then it presents a debatable case of onset weight in complex onsets as discussed previously.

Eastern Popoloca (Kalstrom and Pike 1968, Hajek and Goedemans in prep.) has a stress system where the lexically determined stressed syllable needs to contain a long segment, either a vowel or a consonant, e.g. nčís:sə 'mesh bag' vs. nčís:é 'clay pitcher' or thá:ko 'is teaching' vs. thak:ó 'early in the evening'. What these data suggest is that the stressed syllable in Popoloca is heavy (presumably due to Stress-to-Weight (STW)), a property that can be achieved either by V or C-lengthening. If lengthening is merely



seen as the addition of a mora, and since in C-lengthening, the vowel that *follows* the geminate is the one that gets stressed, then it is reasonable to assume that the medial geminates are tautosyllabic moraic onsets (cf. Marshallese §6.4.2).

In Tümpisa Shoshone, Gordon (2005) claims that primary stress docks on the first syllable, unless the first syllable is CV and the second CVV, in which case it shifts to the second syllable. CVC syllables count as light. More importantly, stress optionally shifts to the second syllable if this has a voiceless onset and a short vowel. While this pattern is entirely compatible with what has been presented in this thesis, the original source, i.e. Dayley (1989: 440), shows that things are not as straightforward, particularly given that there is large variation, e.g. [mə.sú.ru.hjɪn.na] or [mə.su.rù.hjɪn.nà] or [mə.su.rù.hjɪn.na] ‘to rub’. It is also unclear why stress should shift to the peninitial in a word like: [kút.tjɪn.na]~[kut.tjɪn.na] ‘to hit’ given that the initial also has a voiceless onset. As it stands and without any additional requirements, such misalignment of stress is unwarranted, since both initial and peninitial should be equally heavy and thus the leftmost one should receive stress due to better alignment. Furthermore, there is also a case where optionality should not arise in the first place, e.g. [ut.tjɪn.na] or [út.tjɪn.na] ‘to give’. Under the assumption that a voiceless onset renders its syllable heavy, stress here would need to consistently arise on the second syllable. In sum, the Shoshone data do not as yet adequately support onset moraicity.

Even more obscure are cases like the ones Rialland (1993) mentions<sup>14</sup>, where apparently compensatory lengthening occurs after the loss of some onsets in Tyrone Irish, Gyore Moore and the Westphalian dialect of Soest. The data are quite unclear, which is why I do not discuss them in any detail here. The interested reader may consult Rialland and the original sources for further discussion.

Finally, in Gadsup, Frantz and Frantz (1966) devote a single line to state that among other factors, the quality of an onset, i.e. stop vs. non-stop matters for stress with the former being more stress-attracting. While at first glance this pattern seems to be different from what has been suggested here, where the relevant property was [voice], I would not rush to any conclusion yet in the absence of relevant examples and fuller discussion.

What these data then highlight along with the ones in Chapter 7, is that there is still a lot to be understood about all these languages, but also that future documentations

---

<sup>14</sup> I have been unable to trace a dialect by the name of Gyore Moore. Perhaps, Gyore is the French rendering of the Moore dialect Zaore or Joore, reported in the Ethnologue and retrieved on 12 March 2006: [http://www.ethnologue.com/14/show\\_language.asp?code=MHM](http://www.ethnologue.com/14/show_language.asp?code=MHM).

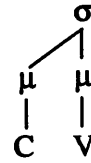
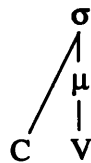
need to pay more attention to the role of onsets, as this can be instrumental in a language's prosodic phonology.

The present thesis has endeavoured to bring the role of the onset into the limelight and unlike many previous theories of stress and weight (Hyman 1985, Hayes 1989, Morén 1999/2001), has proposed that most onset effects are genuine weight effects. Drawing on analogies with tone, I have claimed that the pitch perturbation caused by the voicing of consonants conditions the type of attested moraic onsets, always favouring voiceless over the voiced ones, unless onsets are underlyingly moraic, i.e. are geminates, in which case no such restrictions apply. Consequently, I have proposed that alongside weightless [CVC] and weightful [CVC<sup>μ</sup>] codas, which are imposed on a language-specific basis, the same distinction holds for onsets too, i.e. [CV] and [C<sup>μ</sup>V]. The updated syllabification and weight model for onsets then is:

(22) *Non-moraic onsets*

and

*Moraic onsets*



## Appendix A: Full tableaux for Bella Coola [Ch. 4]

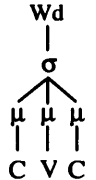
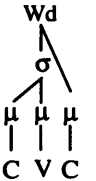
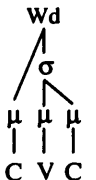
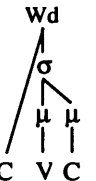
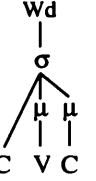
*A clarification in terms of numbering of the examples in the appendix:*

Numbering of the tableaux follows the same number used in the main text so as to facilitate cross-checking. A few ranking arguments included in the Hasse diagrams of Ch.4 are extracted from the full versions of the tableaux that are presented here, e.g. \*NUC/OBSTR >> SLH. Note that DEP- $\mu$  violations in brackets are those which are incurred by a non-moraically specified input. MAX- $\mu$  violations in brackets are those which are incurred by a moraically specified input. It should be clear that even if these are counted, the outcome stays the same.

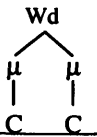
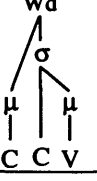
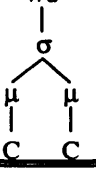
(20) Full ranking for CVC --- moraic input

$/C_{(\mu)}V_{\mu}C_{(\mu)}/ \text{ --- } [[CV_{\mu}C_{\mu}]_{\sigma}]_{PrWd}$

MPARSE >> ONSET, DEP- $\mu$  >> \*MORAIC ONSET >> MAX- $\mu$

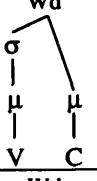
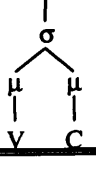
	MPARSE	ONSET	DEP- $\mu$	*MORAIC ONSET	MAX- $\mu$
a. 			(*)(*)	*!	
b. 			(*)(*)	*!	
c. 		*!	(*)(*)		
d. 		*!	(*)		(*)
e. $\emptyset$	*!				
f. 			(*)		(*)

(35) /C<sub>(μ)</sub>C<sub>(μ)</sub>/ --- [C<sub>μ</sub>C<sub>μ</sub>]<sub>PrWd</sub>: \*NUC/OBSTR, DEP-SEG >> MPARSE >> DEP-μ, SLH

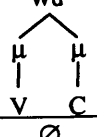

	*NUC/ OBSTR	DEP-SEG	MPARSE	DEP-μ	SLH
a. 				(*)(*)	*
b. ∅			*!		
c. 		*!		(*)*	
d. 	*!			(*)(*)	

Note that things are in fact more complex. If the considered input is /C<sub>μ</sub>C<sub>μ</sub>/ and if DEP-μ is understood in terms of mora number preservation, then (c) does not violate it. Apparently, a more finely-grained version of this is needed, e.g. DEP-LINK-μ (Morén 1999/2001) which militates against the insertion of moras linked to particular segments, here for instance (c) violates DEP-LINK-μ by means of the mora inserted by the vowel.

(39) /V<sub>μ</sub>C<sub>μ</sub>/ --- [[V<sub>μ</sub>C<sub>μ</sub>]<sub>σ</sub>]<sub>PrWd</sub>: PARSE-SEG >> NO CODA

	PARSE-SEG	NO CODA
a. 	*!	
b. 		*

(40) /V<sub>μ</sub>C<sub>(μ)</sub>/ --- [[V<sub>μ</sub>C<sub>μ</sub>]<sub>σ</sub>]<sub>PrWd</sub>: MORAIC CODA, MPARSE >> DEP-μ, SLH >> NO CODA

	MORAIC CODA	MPARSE	DEP-μ	SLH	NO CODA
a. 			(*)	*!	
b. ∅		*!			
c. 	*!				

<p>d.</p> <pre>       Wd               σ      / \     μ   μ               V   C           </pre>				(*)	*
--	--	--	--	-----	---

- (45) /CV<sub>μ</sub>CC/ --- [[CV<sub>μ</sub>C<sub>μ</sub>]<sub>σ</sub>C<sub>μ</sub>]<sub>PrWd</sub>  
 \*CC]<sub>CODA</sub>, MLIC >> DEP-μ, ONS >> \*MORAIC ONS >> MAX-μ

CV <sub>μ</sub> CC	*CC] <sub>CODA</sub>	MLIC	DEP-μ	ONS	*μ ONS	MAX-μ
<p>a.</p> <pre>       Wd               σ      /   \     μ μ μ               C V C C           </pre>			**			
<p>b.</p> <pre>       Wd               σ      /     \     μ μ μ μ                 C V C C           </pre>			***!		*	
<p>c.</p> <pre>       Wd               σ      /   \     μ μ μ               C V C C           </pre>			***!	*		
<p>d.</p> <pre>       Wd               σ      /   \     μ μ μ               C V C C           </pre>		*!	*			
<p>e.</p> <pre>       Wd               σ      /   \     μ μ μ               C V C C           </pre>		*!	**	*		
<p>f.</p> <pre>       Wd               σ      /   \     μ μ μ               C V C C           </pre>	*!		**			

Had the input been fully moraic, then the result would be identical. DEP-μ would not be violated at all, so candidates (b)-(f) would be ruled out by virtue of their markedness violations. Candidate (a) would incur just one violation of lowest-ranked MAX-μ without affecting the outcome in any way.

## Appendix B: Bislama patterns according to Camden (1977) [Ch. 7]

Note that the examples from Camden do *not* imply that the words supplied arise with the stress patterns he offers. This is why a stress mark is missing from words where there is no explicit reference to stress. Overt reference of stress only appears in Lynch (1975). Examples come from Camden (1977; henceforth Ca), Lynch (1975; henceforth L) and Crowley (1995; henceforth Cr). Often, no exact examples could be found in which case, the closest examples to the required pattern are offered.

- (26) two-syllable words: penult stress: *kala* 'colour' (L: 193), *loto* 'motor vehicle' (L: 193)  
 (27) three-syllable words: penult stress: *kampani* (Ca: 44) 'group', *olketa* (L: 194) 'all' But final stress if:

i) final  $\sigma$  = CVC: *hamarem* (Ca: 34) 'to hammer', *bloblokem* 'to hinder, oppose one another'

▪ except in:  $\sigma$ .(C)(C)(C)VC.CVC; some examples include:

- *o.pen.tin* (Ca: 79) 'tin opener', *a.gen.sem* (Cr: 33) 'against', *a.res.tem* (Cr: 36) 'to arrest', *han.dred.fut* (Ca: 34) 'a centipede' CVC.CCVC.CVC [no RED], *skra.skra.sem* (Ca: 105) 'to grate', *brod.kas.tem* (Ca: 14) 'to broadcast', *gi.vim.bak* (Ca: 28) 'to return', *ho.lem.gud* (Ca: 38) 'to hold carefully', *ka.nel.pet* (Ca: 44) 'can't help it (phrase)'

ii) final  $\sigma$  = CCV: *presbitri* (Ca: 86) 'one of the five regional Presbyteries'

▪ except in:

- a) V. 'CVC.CCV: *a.sem.bli* (Cr: 37) 'Presbyterian church assembly meeting', *me.lek.tri* (Ca: 66) CV.CVC.CCV 'a tree with white sap'

iii) final  $\sigma$  = CCVC: *pasenfrut* (Ca: 81) 'passionfruit vine', *graenston* (Ca: 30) 'a grinding wheel for sharpening tools, axes, etc.', *atebrin* (Cr: 37), *bi.sne.sman* (Cr: 47), *springmatres* (Cr: 230) 'inner spring mattress', *trabolples* (Cr: 249) 'place where life is troubled, difficult'

Other exceptions in trisyllabic words involve:

- (28) V.CV.'VC: *o.ra.et* (Ca: 79), *ol.ba.ot* (Ca: 77) 'not in the accepted way (adv.)'  
 (29) CVC.V.'VC [or CV.CV.'VC]: *go.da.on* (Ca: 29) 'go down', *go.ra.on* (Ca: 30) 'go round', *ha.re.ap* (Ca: 35) 'hurry up', *pe.ma.ot* (Ca: 82) 'to buy'  
 (30) CCVC.V.'CV [or CCV.CV.'CV]: [no exact example], cf. *kle.va.man* (Ca: 49) 'person able to recognise sorcery or help people in trouble'  
 (31) four syllable words:  $\sigma\sigma\sigma\sigma$   
 but final stress if  $\sigma\sigma\sigma$ . 'CVC, e.g. *alumijam* (Ca: 2), *enkarajem* (Ca: 20) 'to encourage'; except penult and ultima are: 'CVC.CVC  
 (32) five syllable words:  $\sigma\sigma\sigma\sigma\sigma$ , but final stress if  $\sigma\sigma\sigma\sigma$ . 'CVC: e.g. *lae(k)laekem* (Ca: 53) 'to like/love one another', *purumpurumbut* ? (Ca: 87, ...CVC.CVC) 'to dance with a heavy stamping action'  
 (33) six syllable words:  $\sigma\sigma\sigma\sigma\sigma\sigma$ , e.g. *naranarafala* (Ca: 72) 'differing, diverse'

## References

- Abo, Takaji, Byron W. Bender, Alfred Capelle and Tony DeBrum (1976) *Marshallese-English dictionary*. Honolulu: University Press of Hawaii
- Abrahamson, Arno and Joyce Abrahamson (1984) Os fonemas da língua Júma. In *Estudos sobre línguas tupí do brasil [Série Lingüística 11]*, R. Dooley (ed.), pp. 157-174, Brasília: Summer Institute of Linguistics
- Abramson, Arthur (1986) The perception of word-initial consonant length: Pattani Malay. *Journal of the International Phonetic Association* 16: 8-16
- Abramson, Arthur (1987) Word-initial consonant length in Pattani Malay. *Proceedings of the XIth International Congress of Phonetic Sciences*, Tallinn, pp. 68-70
- Abramson, Arthur (1991) Amplitude as a cue to word-initial consonant length: Pattani Malay. *Proceedings of the XIIth International Congress of Phonetic Sciences*, Aix-en-Provence, pp. 98-101
- Abramson, Arthur (1999) Fundamental frequency as a cue to word-initial length: Pattani Malay. *Proceedings of the XIVth International Congress of Phonetic Sciences*, San Francisco, pp. 591-594
- Afroudakis, Angelos [Αφρουδάκης, Άγγελος] (1985) *Collection of language material from the island of Samothraki [Συλλογή γλωσσικού υλικού από τη νήσο Σαμοθράκη]*. Manuscript No. 1293, Research Centre of Modern Greek Dialects and Idioms, Academy of Athens [Κέντρον Ερεύνης των Νεοελληνικών Διαλέκτων και Ιδιωμάτων, Ι.Α.Ν.Ε Ακαδημίας Αθηνών]
- Akinlabi, Akinbiyi (1993) Underspecification and the Phonology of Yoruba /r/. *Linguistic Inquiry* 24(1): 139-160
- Al-Harbi, Awwad (2005) Nankina Trochees. ms., Umm Al-Qura University. Available as ROA-724
- Albright, Adam (2002) Base selection in analogical change in Yiddish. In *The Proceedings of the 28th Annual Meeting of the Berkeley Linguistics Society (BLS 28)*, pp. 1-13
- Alderete, John (1999) *Morphologically Governed Accent in Optimality Theory*. Doctoral Dissertation, Amherst: University of Massachusetts, Published 2001. New York and London: Routledge

- Alderete, John, Jill Beckman, Laura Benua, Amalia Gnanadesikan, John McCarthy and Suzanne Urbanczyk (1999) Reduplication with fixed segmentism. *Linguistic Inquiry* 30: 327-364
- Andriotis, Nikolaos [Ανδριώτης, Νικόλαος] (1926-1928) *Collection of language material from Samothraki* [Συλλογή γλωσσικού υλικού εκ Σαμοθράκης]. Manuscript No: 449, Research Centre of Modern Greek Dialects and Idioms, Academy of Athens [Κέντρον Ερεύνης των Νεοελληνικών Διαλέκτων και Ιδιωμάτων, Ι.Α.Ν.Ε Ακαδημίας Αθηνών]
- Anttila, Arto (1994) Consonant Gradation is Epenthesis. Handout from paper presented at the Stanford Phonology Workshop (Feb 1994)
- Anttila, Arto (1997) Deriving variation from grammar: a study of Finnish genitives. In *Variation, change and phonological theory*, Frans Hinskens, Roeland van Hout and Leo Wetzels (eds.), pp. 35-68, [also available as ROA-63], Amsterdam: John Benjamins
- Arvaniti, Amalia and Sharon Rose (2003) Two sources of evidence against moraic timing of geminates. Handout from the LSA Meeting, Atlanta, GA, January 2003
- Ashton, E. O., E. M. K. Mulira, E. G. M. Ndawula and A. N. Tucker (1954) *A Luganda Grammar*. London: Longmans, Green and Co
- Baertsch, Karen and Stuart Davis (2003) The Split Margin Approach to Syllable Structure. *ZAS Working Papers in Linguistics* 32: 1-14
- Bagemihl, Bruce (1991) Syllable structure in Bella Coola. *Linguistic Inquiry* 22: 589-646
- Bagemihl, Bruce (1998) Maximality in Bella Coola (Nuxalk). In *Salish Languages and Linguistics, Theoretical and Descriptive Perspectives*, Ewa Czaykowska-Higgins and M. Dale Kinkade (eds.), pp. 71-98, Berlin, New York: Mouton de Gruyter
- Bakovic, Eric (1996) Foot harmony and quantitative adjustments. ms., Rutgers University. Available as ROA-168
- Bao, Zhiming (1990) *On the Nature of Tone*. Doctoral Dissertation, MIT
- Beckman, Jill (1999) *Positional faithfulness: an optimality theoretic treatment of phonological asymmetries*. Doctoral Dissertation 1998, University of Massachusetts, Amherst. Published 1999, New York, London: Garland
- Bell, Alan (1978) Syllabic consonants. In *Universals of Human Language, vol. 2: Phonology*, Joseph H. Greenberg (ed.), pp. 153-201, Stanford, CA: Stanford University Press
- Bender, Byron W. (1976) *Spoken Marshallese*. Honolulu: University Press of Hawaii
- Bender, Byron W. (1991) On the category distributive. In *Currents in Pacific Linguistics: papers on Austronesian languages and ethnolinguistics in honour of George W. Grace*, Robert Blust (ed.), pp. 11-26, Canberra: Department of Linguistics, Australian National University



- Benua, Laura (1995) Identity effects in morphological truncation. In *Papers in Optimality Theory*, Jill Beckman, Laura Walsh Dickey and Suzanne Urbanczyk (eds.), pp. 77-136, Amherst, MA: GLSA
- Bermúdez-Otero, Ricardo (2001) Underlyingly nonmoraic coda consonants, faithfulness and sympathy. ms., University of Manchester
- Blevins, Juliette (2001) Where have all the onsets gone? Initial consonant loss in Australian Aboriginal languages. In *Forty years on: Ken Hale and Australian languages*, Jane Simpson, David Nash, Mary Laughren, Peter Austin, Barry Alpher (eds.), pp. 481-492, Canberra: Pacific Linguistics
- Blumenfeld, Lev (2005a) Review of *Phonological Augmentation in Prominent Positions*, by Jennifer Smith. Review for *The Linguist List*, <http://linguistlist.org/>, available as LL Issue: 16.2194
- Blumenfeld, Lev (2005b) Stress and lenition in Karo. ms., Stanford University
- Blumenfeld, Lev (2005c) Too many solutions: how prosody and segments interact. Handout from talk presented at *The Annual Meeting of the Chicago Linguistics Society, CLS 41*, April 2005
- Booij, Geert (1995) *The phonology of Dutch*. Oxford: Clarendon Press
- Borrelli, Doris (2002) *Raddoppiamento Sintattico in Italian: A Synchronic and Diachronic Cross-Dialectical Study*. New York and London: Routledge
- Bradshaw, Mary (1999) *A Crosslinguistic Study of Consonant-Tone Interaction*. Doctoral Dissertation, Ohio State University
- Breen, Gavan (1992) Some problems in Kukatj phonology. *Australian Journal of Linguistics* 12(1): 1-43
- Breen, Gavan and Rob Pensalfini (1999) Arrernte: A language with no syllable onsets. *Linguistic Inquiry* 30: 1-25
- Broselow, Ellen, Su-I Chen and Marie Huffman (1997) Syllable weight: convergence of phonology and phonetics. *Phonology* 14: 47-82
- Buller, Barbara, Ernest Buller and Daniel Everett (1993) Stress Placement, Syllable Structure, and Minimality in Banawá. *International Journal of American Linguistics* 59: 280-293
- Camden, Pastor Bill (1977) *A descriptive dictionary: Bislama to English*. Rosebury, Australia: Bridge Printery
- Capell, Arthur (1962) *Some Linguistic Types in Australia*. Oceanic Linguistics Monographs, Number 7, Sydney: University of Sydney

- Carlson, Robert J. (1990) *A grammar of Supyire: Kampwo dialect*. Doctoral Dissertation, University of Oregon
- Casali, Roderic (1996) *Resolving hiatus*. Doctoral Dissertation, UCLA
- Clements, George N. (1986) Compensatory Lengthening and Consonant Gemination in LuGanda. In *Studies in Compensatory Lengthening*, Leo Wetzels and Engin Sezer (eds.), pp. 37-77, Dordrecht: Foris
- Clements, George N. (1990) The role of the sonority cycle in core syllabification. In *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, John Kingston and Mary E. Beckman (eds.), pp. 283-333, Cambridge: Cambridge University Press
- Clements, George N. and Samuel Jay Keyser (1983) *CV Phonology: A Generative Theory of the Syllable*. Cambridge, MA: MIT Press
- Cohn, Abigail (2003) Phonological structure and phonetic duration: The role of the mora. *Working Papers of the Cornell Phonetics Laboratory* 15: 69-100
- Coleman, John S. (1996) Declarative Syllabification in Tashlhiyt Berber. In *Current Trends in Phonology: Models and Methods*, Volume 1, Jacques Durand and Bernard Laks (eds.), pp. 177-218, Salford: European Studies Research Institute
- Connell, Bruce A., Richard A. Hayward and John Abraha Ashkaba (2000) Observations on Kunama tone. *Studies in African Linguistics* 29: 1-42
- Crosswhite, Katherine (2001) Predicting syllabicity and moraicity in Dihovo Macedonian. In *University of Rochester Working Papers in the Language Sciences* 2(1), J. S. Magnuson and K. M. Crosswhite (eds.), pp. 1-22
- Crowhurst, Megan and Mark Hewitt (1997) Boolean operations and constraint interactions in Optimality Theory. Unpublished manuscript. Chapel Hill, NC: University of North Carolina; Waltham, MA: Brandeis University. Available as ROA-229
- Crowley, Terry (1995) *A new Bislama dictionary*. Suva, Fiji: Institute of Pacific Studies, Pacific Languages Unit, Vila, Vanuatu
- Crowley, Terry (2004) *Bislama Reference Grammar*. Oceanic Linguistics Special Publication No. 31, Honolulu: University of Hawai'i Press
- Cruttenden, Alan (2001) *Gimson's Pronunciation of English*. 6th Edition, London: Arnold Publishers
- Curtis, Emily K. J. (2003) *Geminate weight: case studies and formal models*. Doctoral Dissertation, University of Washington

- Davis, Stuart (1985) *Topics in Syllable Geometry*. Doctoral Dissertation, University of Arizona, Tucson
- Davis, Stuart (1988) Syllable onsets as a factor in stress rules. *Phonology* 5: 1-19
- Davis, Stuart (1999) On the representation of initial geminates. *Phonology* 16: 93-104
- Davis, Stuart and Gina Torretta (1998) An optimality-theoretic account of compensatory lengthening and geminate throwback in Trukese. *NELS* 28: 111-125
- Dayley, Jon P. (1989) *Tümpisa (Panamint) Shoshone grammar* [University of California Publications in Linguistics 115]. Berkeley: University of California Press
- de Chene, Brent and Stephen R. Anderson (1979) Compensatory Lengthening. *Language* 55: 505-535
- de Lacy, Paul (to appear) The interaction of tone, sonority and prosody. In *The Cambridge Handbook of Phonology*, Paul de Lacy (ed.), Cambridge: Cambridge University Press
- de Lacy, Paul (1997) Prosodic Categorisation. MA thesis, University of Auckland. Available as ROA-236
- de Lacy, Paul (1999) Tone and prominence. ms., University of Massachusetts, Amherst
- de Lacy, Paul (2000) Markedness in Prominent Positions. In *Proceedings of HUMIT, MIT Working Papers in Linguistics* 40, Ora Matushansky et al (eds.) [also available as ROA-432]
- de Lacy, Paul (2002) The interaction of tone and stress in Optimality Theory. *Phonology* 19: 1-32
- Delgutte, Bertrand (1982) Some Correlates of Phonetic Distinctions at the Level of the Auditory Nerve. In *The Representation of Speech in the Peripheral Auditory System*, R. Carlson and B. Granström (eds.), pp. 131-150, Amsterdam: Elsevier
- Dell, François and Mohamed Elmedlaoui (1985) Syllabic Consonants and Syllabification in Imdlawn Tashlhiyt Berber. *Journal of African Languages and Linguistics* 7: 105-130
- Dell, François and Mohamed Elmedlaoui (1988) Syllabic Consonants in Berber: Some New Evidence. *Journal of African Languages and Linguistics* 10: 1-17
- Donohue, Mark (1999) *A grammar of Tukang Besi*. Grammar Library series No. 20. Berlin: Mouton de Gruyter
- Downing, Laura (1998) On the Prosodic Misalignment of Onsetless Syllables. *Natural Language and Linguistic Theory* 16: 1-52

- Downing, Laura (1999) Verbal reduplication in three Bantu languages. In *The Prosody-Morphology Interface*, René Kager, Harry van der Hulst and Wim Zonneveld (eds.), pp. 62-89, Cambridge: Cambridge University Press
- Duanmu, San (1990) *A formal study of syllable, tone, stress and domain in Chinese languages*. Doctoral dissertation, MIT
- Elbert, Samuel H. (1988) *Echo of a culture: A Grammar of Rennell and Bellona*. Oceanic Linguistics Special Publication, No. 22, Honolulu: University of Hawaii Press
- Elenbaas, Nine and René Kager (1999) Ternary rhythm and the lapse constraint. *Phonology* 16: 273-329
- Evans, Nick (1995) Current issues in the phonology of Australian languages. In *The handbook of phonological theory*, John Goldsmith (ed.), pp. 723-761, Cambridge, MA: Blackwell
- Everett, Daniel (1988) On Metrical Constituent Structure in Pirahã Phonology. *Natural Language & Linguistic Theory* 6: 207-246
- Everett, Daniel (1996) Prosodic Levels and Constraints in Banawá and Suruwahá. Available as ROA-496
- Everett, Daniel (1997) Syllable Integrity. In *Proceedings of WCCFL 16*, CSLI/University of Chicago Press
- Everett, Daniel (2003) Iambic Feet in Paumari and the Theory of Foot Structure. *Linguistic Discovery* 2: 22-44
- Everett, Daniel and Keren Everett (1984) On the Relevance of Syllable Onsets to Stress Placement. *Linguistic Inquiry* 15: 705-711
- Everett, Keren (1998) Acoustic Correlates of Stress in Pirahã. MA Thesis, University of Pittsburgh
- Fabb, Nigel (1997) *Linguistics and literature: language in the verbal arts of the world*. Oxford: Blackwell
- Fanselow, Gisbert and Caroline Féry (2002) Ineffability in Grammar. In *Resolving conflicts in grammars: optimality theory in syntax, morphology, and phonology*, Sonderheft 11. Linguistische Berichte, Gisbert Fanselow and Caroline Féry (eds.), pp. 265-307, Hamburg: Helmut Buske Verlag
- Frantz, Chester I. and Marjorie Frantz (1966) Gadsup phoneme and toneme units. Linguistic Circle of Canberra Publications A 7: 1-11, Papers in New Guinea Linguistics 5, Canberra: Linguistic Circle of Canberra, Australian National University

- Gabas Jr., Nilson (1998) *Estudo fonológico da língua Karo*. München and Newcastle: LINCOM Europa
- Gabas Jr., Nilson (1999) *A Grammar of Karo (Tupi)*. Doctoral dissertation, University of California, Santa Barbara
- Gafos, Diamandis (1998) Eliminating long-distance consonantal spreading. *Natural Language and Linguistic Theory* 16: 223-278
- García-Bellido, Paloma and Beatrice Clayre (1997) Prosodic constraints and representations in the Berawan word. *Oxford Working Papers in Linguistics, Philology, and Phonetics* 2: 17-52
- Garrett, Edward (1999) Minimal Words Aren't Minimal Feet. In *UCLA Working Papers in Linguistics, vol. 1, Papers in Phonology* 2, Matthew Gordon (ed.), pp. 68-105
- Goedemans, Rob (1996) An Optimality Account of Onset-Sensitive Stress in Quantity-Insensitive Languages. *Linguistic Review* 13: 33-47
- Goedemans, Rob (1997) Putting the Mathimathi stress rule in its proper perspective. *Australian Journal of Linguistics* 17: 43-69
- Goedemans, Rob (1998) *Weightless segments*. The Hague: Holland Academic Graphics
- González, Carolina (2003) *The effect of stress and foot structure on consonantal processes*. Doctoral Dissertation, University of Southern California
- Goodenough, Ward H. and Hiroshi Sugita (1980) *Trukese-English Dictionary*. Philadelphia: The American Philosophical Society
- Gordon, Matthew (1998) A fortition-based approach to Balto-Fennic-Sámi consonant gradation. *Folia Linguistica Historica* 18: 49-79
- Gordon, Matthew (1999) *Syllable Weight: Phonetics, Phonology, and Typology*. Doctoral Dissertation, UCLA
- Gordon, Matthew (2002) A phonetically-driven account of syllable weight. *Language* 78: 51-80
- Gordon, Matthew (2004) Syllable weight. In *Phonetically Based Phonology*, Bruce Hayes, Robert Kirchner, and Donca Steriade (eds.), pp. 277-312, Cambridge: Cambridge University Press
- Gordon, Matthew (2005) A perceptually-driven account of onset-sensitive stress. *Natural Language and Linguistic Theory* 23: 595-653
- Green, Antony Dubach (2002) Word, foot, and syllable structure in Burmese. ms., University of Potsdam, available as ROA-551

- Green, Thomas (1995) The stress window in Pirahã: a reanalysis of rhythm in Optimality Theory. ms., MIT, available as ROA-45
- Gussenhoven, Carlos (to appear) Vowel duration, syllable quantity and stress in Dutch. In *The Nature of the Word. Essays in Honor of Paul Kiparsky*, K. Hanson and S. Inkelas (eds.), Cambridge, MA: MIT Press
- Gussenhoven, Carlos and Haïke Jacobs (2005) *Understanding Phonology*. First edition 1998, London: Hodder Arnold
- Hajek, John and Rob Goedemans (in prep.) On the representation of onset weight. ms., University of Melbourne, Utrecht University and Universiteit Leiden
- Hajek, John and Rob Goedemans (2003) Word-initial geminates and stress in Pattani Malay. *The Linguistic Review* 20: 79-94
- Hale, Ken (1981) Part III: Damin Vocabulary. In *A Preliminary dictionary of Lardil*, Ken Hale et al., pp. 294-299, Document 0028 of the Aboriginal Studies Electronic Data Archive (ASEDA), Canberra: AIATSIS (Australian Institute of Aboriginal and Torres Strait Islander Studies)
- Hale, Ken and David Nash (1997) Damin and Lardil Phonotactics. In *Boundary rider: essays in honour of Geoffrey O'Grady*, Darrell Tryon and Michael Walsh (eds.), Pacific Linguistics C-136, pp. 247-259, Canberra: Research School of Pacific and Asian Studies, the Australia National University
- Hall, Tracy A. (2000) Syllabically conditioned coalescence and deletion in Zoque: an optimality-theoretic approach. *Linguistics* 38 (4): 711-738
- Halle, Morris and Samuel Jay Kayser (1971) *English Stress: Its Form, its Growth, and its Role in Verse*. New York: Harper & Row
- Halle, Morris and Kenneth Stevens (1971) A note on laryngeal features. *RLE Quarterly Progress Report* 101: 198-312
- Ham, William (2001) *Phonetic and phonological aspects of geminate timing*. New York and London: Routledge
- Hammond, Michael (1999) *The Phonology of English*. Oxford: Oxford University Press
- Hargus, Sharon (2001) Quality sensitive stress reconsidered. *University of Washington Working Papers in Linguistics* 20: 25-56
- Harrison, Sheldon P. (1976) *Mokilese Reference Grammar*. Honolulu: University of Hawaii Press

- Hayes, Bruce (1989) Compensatory Lengthening in Moraic Phonology. *Linguistic Inquiry* 20: 253-306
- Hayes, Bruce (1995) *Metrical Stress Theory: Principles and Case Studies*. Chicago: University of Chicago Press
- Heisenberg, August (1921) Die Liquida P im Dialekt von Samothrake. In *Αφιέρωμα εις Γ.Ν. Χατζιδάκι* [For G.N Chatzidakis], pp. 89-99, Athens: P.D. Sakellariou
- Hendricks, Sean (1999) *Reduplication without template constraints: A study in bare-consonant reduplication*. Doctoral Dissertation, University of Arizona
- Hendricks, Sean (2001) Bare-Consonant Reduplication without Prosodic Templates: Expressive Reduplication in Semai. *Journal of East Asian Linguistics* 10: 287-306
- Hermans, Ben (2001) CL in OT. *Linguistics in the Netherlands* 18: 139-150
- Hollenbach, Barbara (1977) Phonetic vs. phonemic correspondence in two Trique dialects. In *Studies in Otomanguean phonology*, William R. Merrifield (ed.), pp. 35-67, Dallas: Summer Institute of Linguistics and University of Texas at Arlington
- Hubbard, Kathleen (1994) *Duration in moraic theory*. Doctoral Dissertation, University of California, Berkeley
- Hulst, Harry van der (1984) *Syllable Structure and Stress in Dutch*. Dordrecht: Foris
- Hyde, Brett (2001) *Metrical and Prosodic Structure in Optimality Theory*. Doctoral Dissertation, Rutgers University
- Hyman, Larry (1985) *A Theory of Phonological Weight*. Publications in Language Sciences 19 Dordrecht: Foris [reissued in 2003 by the University of Chicago Press]
- Itô, Junko and Armin Mester (1986) The Phonology of Voicing in Japanese: Theoretical Consequences for Morphological Accessibility. *Linguistic Inquiry* 17: 49-73
- Itô, Junko and Armin Mester (1992) Weak Layering and Word Binarity. ms., University of California, Santa Cruz
- Itô, Junko and Armin Mester (1994) Reflections on CodaCond and Alignment. In *Phonology at Santa Cruz [PASC]* 3, J. Merchant, J. Padgett, and R. Walker (eds.), pp. 27-46, [available as ROA-141], Linguistics Research Center, University of California, Santa Cruz
- Jakobson, Roman (1931) Die Betonung und ihre Rolle in der Wort- und Syntagma-phonologie. *Travaux du cercle linguistique de Prague* 4. Reprinted in R. Jakobson (1962) *Selected Writings I: Phonological Studies*, pp. 117-136, Mouton: The Hague

- Jensen, Christian (2004) *Stress and Accent: Prominence relations in Southern Standard British English*. Doctoral Dissertation, University of Copenhagen
- Jones, Daniel (1909) *The pronunciation of English*. 4th edn. 1956, Cambridge: Cambridge University Press
- Kager, René (1989) *A Metrical Theory of Stress and Destressing in English and Dutch*. Dordrecht: Foris
- Kager, René (1999) *Optimality Theory*. Cambridge: Cambridge University Press
- Kalstrom, Marjorie K. and Eunice V. Pike (1968) Stress in the Phonological System of Eastern Popoloca. *Phonetica* 18: 16-30
- Katamba, Francis and Peter Cooke (1987) Ssematimba ne Kikwabanga: the music and poetry of a Ganda historical song. *World of Music* 29(2): 49-68
- Katsanis, Nikolaos [Κατσάνης, Νικόλαος] (1996) *The dialect of Samothraki Greek [Το γλωσσικό ιδίωμα της Σαμοθράκης]*. Municipality of Samothraki [Δήμος Σαμοθράκης]
- Kavitskaya, Darya (2002) *Compensatory Lengthening: Phonetics, Phonology, Diachrony*. New York and London: Routledge
- Kenstowicz, Michael (1994a) *Phonology in Generative Grammar*. Oxford: Blackwell
- Kenstowicz, Michael (1994b) Sonority-Driven Stress. Available as ROA-33
- Kingston, John and David Solnit (1988a) The Inadequacy of Underspecification. In *Proceedings of NELS 19*. Amherst, MA: GLSA
- Kingston, John and David Solnit (1988b) The tones of consonants. ms., University of Massachusetts, Amherst and University of Michigan
- Kiparsky, Paul (to appear) Fenno-Swedish Quantity: Contrast in Stratal OT. In Bert Vaux (ed.); ms. available at: <http://www.stanford.edu/~kiparsky/>
- Kiparsky, Paul (2000) Opacity and Cyclicity. *The Linguistic Review* 17: 351-367
- Kiparsky, Paul (2002) Syllables and Moras in Arabic. In *The Syllable in Optimality Theory*, Caroline Féry and Ruben van de Vijver (eds.), Cambridge: Cambridge University Press
- Kisseberth, Charles (1970) Vowel elision in Tonkawa and derivational constraints. In *Studies Presented to Robert B. Lees by his Students*, Jerrold M. Sadock and Anthony L. Vanek (eds.), pp. 109-137, Edmonton, AB and Champaign, IL: Linguistic Research
- Kraehenmann, Astrid (2001) Swiss German stops: geminates all over the word. *Phonology* 18: 109-145



- Krekoukias, Dimitrios A. [Κρεκούκας, Δημήτριος Α.] (1964) *Collection of language material from Samothraki* [Συλλογή γλωσσικού υλικού εκ Σαμοθράκης]. Manuscript No 839, Research Centre of Modern Greek Dialects and Idioms, Academy of Athens [Κέντρον Ερεύνης των Νεοελληνικών Διαλέκτων και Ιδιωμάτων, Ι.Α.Ν.Ε Ακαδημίας Αθηνών]
- Ladefoged, Peter and Ian Maddieson (1996) *The sounds of the world's languages*. Oxford; Cambridge, MA: Blackwell
- Lahiri, Aditi and Jacques Koreman (1988) Syllable weight and quantity in Dutch. In *Proceedings of WCCFL 7*, Hagit Borer (ed.), pp. 217-228, Stanford, CA: The Center for the Study of Language and Information, Ventura Hall, Stanford University
- Lavoie, Lisa M. (2001) *Consonant Strength: Phonological Patterns and Phonetic Manifestations*. New York and London: Garland Publishing
- Leben, William (2001) Tonal Feet. Paper presented in Typology of African Prosodic Systems Workshop, Bielefeld University, Germany, May 18-20, 2001
- Lee, Byung-Gun (1996) The Emergence of the Faithful. Available as ROA-166
- Levelt, Clara and Ruben Van de Vijver (1998) Syllable types in cross-linguistic and developmental grammars. In *Fixing Priorities: Constraints in Phonological Acquisition*, René Kager, Joe Pater and Wim Zonneveld (eds.), [also available as ROA-265], Cambridge: Cambridge University Press
- Levin, Juliette (1985) *A Metrical Theory of Syllabicity*. Doctoral Dissertation, MIT
- Lin, Yen-Hwei (1997) Syllabic and moraic structures in Piro. *Phonology* 14: 403-436
- Lin, Yen-Hwei (1998) On Minor Syllables. In *Proceedings of LP'98*, O. Fujimura, B.D. Joseph and B. Palek (eds.), pp. 163-183, Charles University in Prague: The Karolinum Press
- Lombardi, Linda (1990) The nonlinear organization of the affricate. *Natural Language and Linguistic Theory* 8: 375-425
- Lombardi, Linda (1994) *Laryngeal Features and Laryngeal Neutralization*. New York and London: Garland Publishing
- Lombardi, Linda (2002) Markedness and the typology of epenthetic vowels. In the *Proceedings of Linguistics and Phonetics 2002*; available at: <http://www.adn.nu/~ad31175/lp2002/lp2002main.htm>
- Longacre, Robert E. (1957) Proto-Mixtecan. Publication 5 of the Indiana University Research Center in Anthropology, Folklore, and Linguistics, published as Part 3 of Vol. 23, No. 4 of the *International Journal of American Linguistics*
- Lynch, John (1975) Bislama Phonology and Grammar: A Review Article. *Kivung* 8(2): 186-204

- Maddieson, Ian (1997) Phonetic universals. In *Handbook of Phonetic Sciences*, William J. Hardcastle and John Laver (eds.), pp. 619-639, Oxford: Blackwell
- Madsen, Shannon (2000) Stress patterns in Spanish blends. In *UCLA Working Papers in Linguistics, no.4: Papers in Phonology 4*, Adam Albright and Taehong Cho (eds.)
- Mateus, Maria Helena and Ernesto d'Andrade (2002) *The Phonology of Portuguese*. Oxford: Oxford University Press
- Matisoff, James A. (1973) Tonogenesis in Southeast Asia. In *Consonant Types and Tone, Southern California Occasional Papers in Linguistics, No.1*, Larry Hyman (ed.), pp. 71-95
- Matteson, Esther (1965) *The Piro (Arawakan) language*. Berkeley: University of California Press
- McCarthy, John (1979) *Formal Problems in Semitic Phonology and Morphology*. Doctoral Dissertation, MIT, Cambridge, MA
- McCarthy, John (2002) *A Thematic Guide to Optimality Theory*. Cambridge: Cambridge University Press
- McCarthy, John (2003a) Comparative markedness. *Theoretical Linguistics* 29: 1-51
- McCarthy, John (2003b) OT constraints are categorical. *Phonology* 20: 75-138
- McCarthy, John (2003c) Sympathy, Cumulativity, and the Duke-of-York Gambit. In *The Optimal Syllable*, Caroline Féry and Ruben van de Vijver (eds.), pp. 23-76, Cambridge: Cambridge University Press
- McCarthy, John (2005) Candidate chains. Paper presented at OCP2, Tromsø, Norway
- McCarthy, John and Alan Prince (1986) Prosodic Morphology 1986. Report no. RuCCS-TR-32. New Brunswick, NJ: Rutgers University Center for Cognitive Science
- McCarthy, John and Alan Prince (1993) Generalized alignment. In *Yearbook of Morphology*, Geert Booij and Jaap van Marle (eds.), pp. 79-153, Dordrecht: Kluwer
- McCarthy, John and Alan Prince (1994) The emergence of the unmarked: Optimality in prosodic morphology. In *Proceedings of the North East Linguistic Society 24*, Mercè González (ed.), pp. 333-379, Amherst, MA: GLSA
- McCarthy, John and Alan Prince (1995) Faithfulness and reduplicative identity. In *University of Massachusetts Occasional Papers in Linguistics 18*, Jill Beckman, Laura Walsh Dickey and Suzanne Urbanczyk (eds.), pp. 249-384, Amherst, MA: GLSA
- McCarthy, John and Matthew Wolf (2005) Less than zero: Correspondence and the null output. ms., University of Massachusetts, Amherst [Available as ROA-722]

- McCrary, Cristie (2002) Vowel Length, Consonant Length and *Raddoppiamento-Sintattico*: Segment Duration in Italian Revisited. Paper presented at the 10th Manchester Phonology Meeting, Manchester
- McGarrity, Laura W. (2003) *Constraints on patterns of primary and secondary stress*. Doctoral Dissertation, Indiana University
- Mester, Armin (1990) Patterns of Truncation. *Linguistic Inquiry* 20: 478-485
- Michelson, Karin (1988) *A Comparative Study of Lake-Iroquoian Accent*. Dordrecht: Kluwer
- Mohanan, K. P. (1986) *The Theory of Lexical Phonology*. Dordrecht: Reidel
- Mohanan, Tara (1989) Syllable Structure in Malayalam. *Linguistic Inquiry* 20: 589-625
- Morén, Bruce (1999) *Distinctiveness, Coercion and Sonority: A Unified Theory of Weight*. Doctoral Dissertation, University of Maryland at College Park. Published 2001 by Routledge
- Morén, Bruce (2000) The puzzle of Kashmiri stress: implications for weight theory. *Phonology* 17: 365-396
- Moreton, Elliott and Paul Smolensky (2002) Typological Consequences of Local Constraint Conjunction. In *Proceedings of WCCFL 21*, L. Mikkelsen and C. Potts (eds.), Cambridge, MA: Cascadilla Press
- Morphy, Frances (1983) Djapu, a Yolngu dialect. In *Handbook of Australian Languages Vol. III*, R.M.W Dixon and Barry J. Blake (eds.), pp. 1-188, Amsterdam: John Benjamins
- Muller, Jennifer (1999) A unified mora account of Chuukese. In *Proceedings of WCCFL 18*, S. Bird, A. Carnie, J. Haugen and P. Norquest (eds.) Somerville, MA: Cascadilla Press
- Muller, Jennifer (2001) *The Phonology and Phonetics of Word-Initial Geminates*. Doctoral Dissertation, Ohio State University
- Murray, Robert and Theo Vennemann (1983) Sound change and syllable structure in Germanic phonology. *Language* 59: 514-528
- Nater, Hank F. (1977) *Stem list of the Bella Coola language*. PdR Press Publications of Salish Languages 4, Lisse, Holland: Peter de Ridder Press
- Nater, Hank F. (1979) Bella Coola Phonology. *Lingua* 49: 169-187
- Nater, Hank F. (1984) *The Bella Coola language*. Canada. National Museum of Man. Mercury Series. Ethnology Service Papers 92. Ottawa
- Nelson, Nicole (1998) Right Anchor, Aweigh. ms., Rutgers University [also available as ROA-284]

- Nespor, Marina and Irene Vogel (1986) *Prosodic Phonology*. Dordrecht: Foris
- Newton, Brian (1972) Loss of /r/ in a Modern Greek dialect. *Language* 48: 566-572
- Odden, David (2005) *Introducing Phonology*. Cambridge: Cambridge University Press
- Ohala, John J. (1983) The origin of sound patterns in vocal tract constraints. In *The Production of Speech*, J. MacNeilage (ed.), pp. 189-216, New York: Springer Verlag
- Ohala, John J. (1997) The relation between phonetics and phonology. In *The Handbook of Phonetic Sciences*, W. Hardcastle and J. Laver (eds.), pp. 674-694, Oxford: Blackwell
- Ola, Olanike (1995) Properheadedness and Binariness: Prosodic Words in Yoruba. In *Theoretical Approaches to African Linguistics*, Akinbiyi Akinlabi (ed.), pp. 273-293, New Jersey: Africa World Press Inc.
- Ola, Olanike (1997) *Benue-Congo Prosodic Phonology and Morphology*. Lincom Studies in African Linguistics 15, München, Newcastle: Lincom Europa
- Oostendorp, Marc van (1995) *Vowel Quality and Phonological Projection*. Doctoral Dissertation, Tilburg University, available as ROA-84. Published (2000) as *Phonological Projection: A theory of feature content and prosodic structure*, Berlin, New York: Mouton de Gruyter
- Oostendorp, Marc van (2000) Back to the Base. On the Richness of the Base Hypothesis. ms., Meertens Institute, Amsterdam
- Oostendorp, Marc van (2005) Derived environment effects and consistency of exponence. Talk presented at University College London, Dec 2005
- Orgun, Cemil Orhan and Sharon Inkelas (1992) Turkish prosodic Minimality. Paper presented at the 6th International Conference on Turkish Linguistics, Anadolu University, Eskisehir, Turkey
- Orgun, Cemil Orhan and Ronald L. Sprouse (1999) From MPARSE to CONTROL: Deriving Ungrammaticality. *Phonology* 16: 191-224
- Padgett, Jaye (2002) Constraint Conjunction versus Grounded Constraint Subhierarchies in Optimality Theory. ms., University of California, Santa Cruz
- Papadopoulos, Anthimos [Παπαδόπουλος, Άνθιμος] (1927) *Grammar of Modern Greek Northern Idioms* [Γραμματική των Βορείων Ιδιωμάτων της Νέας Ελληνικής]. Athens: P.D. Sakellarios [Εν Αθήναις: Π.Δ. Σακελλάριος]
- Papazachariou, Dimitris and Periclis Politis (2005) Prosodic markers of a narrative type: the case of Greek TV news reports. Paper presented at the 7th International Conference in Greek Linguistics, York, England, 8-10 September 2005

- Parker, Steven G. (2002) *Quantifying the Sonority Hierarchy*. Doctoral Dissertation, University of Massachusetts, Amherst
- Paula, Ruth Wallace de Garcia (1980) *Língua Kaxuyána: fonologia segmental e afixos nominais*. Rio de Janeiro: Museo Nacional
- Payne, David and Furne Rich (1988) Sensitivity to Onset in Arabela Stress. ms., Información de Campo 743. Yarinacocha, Pucallpa, Peru: Instituto Lingüístico de Verano
- Pearce, Mary (2005) Iambicity without stress in Kera. In *Proceedings of the University of Cambridge Third Postgraduate Conference in Language Research (CamLing 2005)*, G. Chatzidamianos et al. (eds.), pp. 271-278, Cambridge: Cambridge Institute of Language Research
- Peng, Long (1992) *A Unified Theory of Tone-Voice*. Doctoral Dissertation, University of Arizona
- Prince, Alan (1997) Topics in Optimality Theory: TETU, GTH and BC. Handout from LSA Class, retrieved on 17/12/05 from: <http://ling.rutgers.edu/gamma/talks/insthdt4.pdf>
- Prince, Alan and Paul Smolensky (1993) *Optimality Theory: Constraint Interaction in Generative Grammar*. Published 2004, Malden, MA; Oxford, UK; Victoria, Australia: Blackwell
- Revithiadou, Anthi (1998) *Headmost Accent Wins: Head Dominance and Ideal Prosodic Form in Lexical Accent Systems*. Doctoral Dissertation, University of Leiden. Published 1999, The Hague: Holland Academic Graphics
- Rialland, Annie (1993) L'allongement compensatoire: nature et modèles. In *Architecture des représentations phonologiques*, B. Laks et A. Rialland (eds.), pp. 59-92, collection "Sciences du Langage", Editions du CNRS
- Rice, Curt (in press) Optimizing gender. To appear in *Lingua*
- Rice, Keren (1992) On Deriving Sonority: A Structural Account of Sonority Relationships. *Phonology* 9: 61-99
- Rice, Keren and Peter Avery (1989) On the Interaction between Sonorancy and Voicing. *Toronto Working Papers in Linguistics* 10: 65-82
- Rich, Furne (1963) Arabela Phonemes. In *Studies in Peruvian Indian Languages I*, Benjamin Elson (ed.), pp. 193-206, Summer Institute of Linguistics Publications in Linguistics and Related Fields 9, Norman, OK: Summer Institute of Linguistics of the University of Oklahoma

- Robinson, Lila W. (1975) Some Phonological Rules of Iowa-Oto (Siouan). *Proceedings of the Mid-America Linguistics Conference*, pp. 439-448, Linguistics Department, University of Kansas, Lawrence
- Roca, Iggy and Wyn Johnson (1999) *A Course in Phonology*. Oxford: Blackwell
- Rosenthal, Sam (1994) *Vowel/Glide Alternation in a Theory of Constraint Interaction*. Doctoral Dissertation, University of Massachusetts, Amherst. Published 1997, New York and London: Garland Publishing
- Rosenthal, Sam and Harry van der Hulst (1999) Weight-by-position by position. *Natural Language & Linguistic Theory* 17: 499-540
- Rubach, Jerzy (1998) A Slovak argument for the onset-rhyme distinction. *Linguistic Inquiry* 29: 168-179
- Rubach, Jerzy (2002) Against subsegmental glides. *Linguistic Inquiry* 33: 672-687
- Sagey, Elizabeth (1986) *The Representation of Features and Relations in Nonlinear Phonology*. Doctoral Dissertation, MIT
- Sapir, Edward (1965) *A grammar of Diola-Fogny*. Cambridge University Press
- Selkirk, Elisabeth (1980) The Role of Prosodic Categories in English Word Stress. *Linguistic Inquiry* 11: 563-606
- Selkirk, Elisabeth (1984) *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge, MA and London: MIT Press
- Selkirk, Elisabeth (1990) A Two-Root Theory of Length. In *University of Massachusetts Occasional Papers* 14: 123-171, Elaine Dunlap and Jaye Padgett (eds.), Amherst, MA: GLSA
- Shaw, Patricia A. (2002) On the edge: Obstruent clusters in Salish. In *Proceedings of the Workshop on the Structure and Constituency of the Languages of the Americas (WSCLA 7)*, L.Bar-el, L. Watt and I. Wilson (eds.)
- Shaw, Patricia A. (2004) Inside Access: The Prosodic Role of Internal Morphological Constituency. In *The Nature of Word: essays in honor of Paul Kiparsky*, K. Hanson and S. Inkelas (eds.), Cambridge, MA: MIT Press
- Sherzer, Joel (1970) Talking Backwards in Cuna: The Sociological Reality of Phonological Descriptions. *Southwestern Journal of Anthropology* 26: 343-353
- Shryock, Aaron M. (1995) *Investigating Laryngeal Contrasts: An Acoustic Study of the Consonants of Musey*. UCLA Working Papers in Phonetics, no. 89

- Sloan, Kelly (1988) Bare-Consonant Reduplication: Implications for a Prosodic Theory of Reduplication. In *Proceedings of WCCFL 7*, Hagit Borer (ed.), pp. 319-330, Stanford, CA: The Center for the Study of Language and Information, Ventura Hall, Stanford University
- Smith, Jennifer L. (2003) Onset sonority constraints and subsyllabic structure. Submitted to *Phonologica 2002*, John R. Rennison, Friedrich Neubarth, and Markus A. Pöchtrager (eds.); [also available as ROA-608]
- Smith, Jennifer L. (2005) *Phonological Augmentation in Prominent Positions*. New York and London: Routledge
- Smolensky, Paul (1993) Harmony, Markedness, and Phonological Activity. Paper presented at the Rutgers Optimality Workshop 1, October 1993, available as ROA-87
- Spaelti, Philip (1999) *Dimensions of variation in multi-pattern reduplication*. Doctoral Dissertation, University of California Santa Cruz
- Spaulding, Craig and Pat Spaulding (1994) *Phonology and grammar of Nankina* [Data papers on Papua New Guinea languages 41]. Ukarumpa, New Guinea: Summer Institute of Linguistics
- Sprouse, Ronald (1996) Vowels that borrow moras: Geminate and Weight in OT. In *Proceedings of the North Eastern Linguistic Society (NELS) 26* (1995), Kiyomi Kusumoto (ed.), pp. 393-408, Amherst: GLSA
- Sprouse, Ronald (1997) A case for enriched inputs. Handout for paper presented at TREND VII, held May 3, 1997 at UCSC. Available as ROA-193
- Steriade, Donca (1990) Moras and other slots. In *Formal Linguistics Society of Midamerica 1*, D. Meyer and S. Tomioka (eds.) pp. 254-280, Madison: Department of Linguistics, University of Wisconsin
- Stevens, Kenneth (1998) *Acoustic Phonetics*. Cambridge, MA and London: MIT Press
- Strehlow, Theodor G. H. (1944) *Aranda Phonetics and Grammar*. Oceania Monographs 7, Sydney University Press
- Sumner, Meghan (1999) Compensatory lengthening as coalescence: analysis and implications. In *Proceedings of WCCFL 18*, S. Bird, A. Carnie, J. Haugen and P. Norquest (eds.), pp. 532-544, Somerville, MA: Cascadilla Press
- Svantesson, Jan-Olof (1983) *Kammu Phonology and Morphology*. Malmö: CWK Gleerup
- Takahashi, Toyomi (1994) Constraint interaction in Aranda stress. In *UCL Working Papers in Linguistics 6*: 479-507

- Topintzi, Nina (2004) Moraic Onsets and WSP Partition in Pirahã. In *Camling 2004, Proceedings of the Second Postgraduate Conference in Language Research*, Cambridge University, Evangelia Daskalaki et al. (eds), pp. 211-218 Cambridge: Cambridge Institute of Language Research
- Topintzi, Nina (2005a) Putting 'commas' at the right place: A note on crucial non-ranking in OT. In *UCL Working Papers in Linguistics 17*, Mary Pearce and Nina Topintzi (eds.), pp. 83-108
- Topintzi, Nina (2005b) Solving the Samothraki Greek Compensatory Lengthening Puzzle. In the *Proceedings of the 17th International Symposium of Theoretical & Applied Linguistics*, Thessaloniki, Greece, 15-17 April 2005
- Tranel, Bernard (1987) *The sounds of French: an introduction*. Cambridge: Cambridge University Press
- Tranel, Bernard (1991) CVC light syllables, Geminates and Moraic Theory. *Phonology* 8: 291-302
- Trubetzkoy, Nikolay Sergeevich (1939) *Grundzüge der Phonologie*. Göttingen: Vandenhoeck and Ruprecht
- Vennemann, Theo (1988) *Preference laws for syllable structure and the exploration of sound change*. Berlin: Mouton
- Vieimeister, Neal (1980) Adaptation of Masking. In *Psychophysical, Physiological and Behavioural Studies in Hearing, Proceedings of the 5th International Symposium on Hearing*, G. van den Brink and F.A. Bilsen (eds.), pp. 190-198, Delft: Delft University Press
- Welmers, William (1962) The phonology of Kpelle. *Journal of African Languages* 1: 69-93
- Wetzels, Leo and Engin Sezer (eds.) (1985) *Studies in Compensatory Lengthening*. Dordrecht: Foris
- Williamson, Kay (1965) *A grammar of the Kolokuma dialect of Ijo*. Second edition 1969, West African Language Monographs 2, Cambridge: Cambridge University Press
- Woodbury, Hanni (1981) The loss of a phoneme. *International Journal of American Linguistics* 47: 103-120
- Yallop, Colin (1977) *Alyawarra: An Aboriginal Language of Central Australia*. Research and Regional Studies 10, Canberra: Australian Institute of Aboriginal Studies
- Yip, Moira (1980) *The Tonal Phonology of Chinese*. Doctoral Dissertation, MIT



- Yip, Moira (1991) Coronals, Consonant Clusters and the Coda Condition. In *The Special Status of Coronals. Internal and External Evidence*, Caroline Paradis and Jean-François Prunet (eds.), pp. 61-78, San Diego: Academic Press
- Yip, Moira (1992) Sonorant vs. Obstruent Codas: A Prosodic Distinction. ms., University of California, Irvine
- Yip, Moira (1995) Tone in East Asian Languages. In *Handbook of Phonological Theory*, J. Goldsmith (ed.), pp. 476-494, Oxford: Blackwell
- Yip, Moira (2002) *Tone*. Cambridge: Cambridge University Press
- Yip, Moira (2003) Some real and not-so-real consequences of comparative markedness. *Theoretical linguistics* 29.1/2: 53-64
- Yupho, Nawanit (1989) Consonant Clusters and Stress Rules in Pattani Malay. *Mon-Khmer Studies* 15: 125-137
- Zamma, Hideki (1998) English Cluster Simplification and Optimality Theory: A Critique of Lexicon Optimization and Richness of the Base. *Tsukuba English Studies* 17: 1-23
- Zec, Draga (1988) *Sonority Constraints on Prosodic Structure*. Doctoral Dissertation, Stanford University
- Zec, Draga (1995) Sonority constraints on syllable structure. *Phonology* 12: 85-129
- Zewen, François-Xavier Nicolas (1977) *The Marshallese language: A study of its phonology, morphology, and syntax*. Berlin: Verlag von Dietrich Reimer
- Zoll, Cheryl (1998) Positional asymmetries and licensing. ms., MIT. Available as ROA-282